

-2001-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 209/213 (98%), Positives = 212/213 (99%)

5
Query: 1 MTKEYEVEDMSKVAIVTGAGQGIGFAIAKRLHADGFKIGVLDYNEETAQAQAVDKLSPEDA 60
+TK+YEVEDMSKVAIVTGAGQGIGFAIAKRLHADGFKIG+LDYNEETAQAQAVDKLSPEDA
Sbjct: 1 LTKKYEVEDMSKVAIVTGAGQGIGFAIAKRLHADGFKIGILDYNEETAQAQAVDKLSPEDA 60

10
Query: 61 VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITTEEQFEKAFAINVGG 120
VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITTEEQFEKAFAINVGG
Sbjct: 61 VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITTEEQFEKAFAINVGG 120

15
Query: 121 TIWGSQAAQKHRELGHGGKIINATSQAGCEGNPNLTIVYGGTKFAVRGITQTLAKDLASE 180
TIWGSQAAQKHRELGHGGKIINATSQAGCEGNPNLTIVYGGTKFAVRGITQTLAKDLASE
Sbjct: 121 TIWGSQAAQKHRELGHGGKIINATSQAGCEGNPNLTIVYGGTKFAVRGITQTLAKDLASE 180

20
Query: 181 GITVNAYAPGIVKTPMMFDIAHEVGKNAGKDDE 213
GITVNAYAPGIVKTPMMF IAHEVGKNAGKDDE
Sbjct: 181 GITVNAYAPGIVKTPMMFAIAHEVGKNAGKDDE 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1773

25 A DNA sequence (GBSx1880) was identified in *S.agalactiae* <SEQ ID 5513> which encodes the amino acid sequence <SEQ ID 5514>. This protein is predicted to be ATP-dependent DNA helicase. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3735(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB38451 GB:L47709 22.4% identity with Escherichia coli

DNA-damage inducible protein ...; putative [Bacillus subtilis]

Identities = 132/461 (28%), Positives = 231/461 (49%), Gaps = 22/461 (4%)

Query: 21 RKYAVVDLEATGAGPNAS--IIQVGIVIIQGNKIIDSYETDVPNPHESLDEHIVHLTGITD 78

+++ V+D+E TG P IIQ+ V+I+ +I + + +NP++S+ I LTGI++

Sbjct: 4 QRFVVIDVETTGNSPKKGDKIIQIAAVVIENGQITERFSKYINPNKSIPAFIEQLTGISN 63

Query: 79 KQLAKAPDFGQVAHHYQLIEDCIFVAHNVKFDANLLAEQLFLEGCELRTPRI-DTVELS 137

+ + F VA ++QL++ FVAHN+ FD + +L G +L + DTVELS

Sbjct: 64 QMVENEQPFEEVAEEVFQLLDGAYFVAHNIHFDLGFVKYELHKAGFQLPDCEVLDTVELS 123

Query: 138 QVFYPCLEKYSLGALAESLNIELTDAHTAIADARATAQLFIKLIKAKISSLPKEVLETILT 197

++ +P E Y L L+E L + H A +DA T +F+++ K+ LP L+ +

Sbjct: 124 RIVFPGFEGYKLTSELQLRHDQPHRADSDAEVTGLIFLEILEKLRQLPYPTLKQLRR 183

Query: 198 FADNLLFESYLLIEEAYQEADFNVPKEYYFWQGLVLKKEKAVGKPKKLSSDFQ----- 250

+ + + + L++ E Y + +++ +A+ +F

Sbjct: 184 LSQHFISDLTHLDMFINENRHTIEIPGYTRFSSFSVREPAIDVRINEDENFSFEIESWE 243

Query: 251 -----VNALLGMDARPKQVVFADLVKAHFNDQTTTFLEAQPGLGKTYGYLLP--LLDQ 302

++ + G + R Q++ V F ++ +EA PG+GKT GYL+P L +

Sbjct: 244 AGNEKALSELMPGYEKRDGQMMMMREVADAFANREHALIEAPPGIGKTIGYLIPALFAK 303

-2002-

Query: 303 SQKQQIIVSVPTKILQDQIMAKEIKHIQELFHIPCHS--IKGPRNYLKLDAFYKSLQVQD 360
 K+ +I+S + +LQ QI+ K++ +Q+LF P + +KG +YL L F + L +D
 Sbjct: 304 KSKKPVIIISTYSTLLQQQILTKDLPIVQDLFPFVTAAILKGQSHYLCYKFEQVLHEED 363

Query: 361 RNRLINRFKMQLLVWLTTETTTGDLDEIKQKQRLSESYFDQLKHDGE-VTQSSLFYDLDFWK 419
 N K QLLVWLTTET TGD+ E+ + +D+L +D + +S + + F++
 Sbjct: 364 DNYDAVLTKAQLLVWLTTETNTGDVAELNLPSSGKLLWDRLAYDDDSYKRSRSEHVIGFYE 423

Query: 420 RSYDKVAQSQLVIINHAYFL-ERVQDDKDFAGKVLVFDEA 459
 R+ +S LVI NH+ L + K + + DEA
 Sbjct: 424 RAKQIAMRSDLVITNHSLLLTDEGSHKKRLPESGTFIIDEA 464
 Identities = 63/195 (32%), Positives = 88/195 (44%), Gaps = 16/195 (8%)

Query: 629 KVVWIDTSMNPILDLSPSEQYAYETIAKRLQDIMTLKQPT-LVLLTSKQTMFMVSDYLDKWEI 687
 +V I M +I D ++ + A+ ++ + KQP LVL TS + V E+
 Sbjct: 720 QVMIPKEMKSIQDTGQPEFIQDTARYIELMAKEKQPKILVLFVTSMDMLKKVHQ-----EL 774

Query: 688 KH-----LTQD-KNGLAYNVKKRFDRGESNLLLTGSFWEGVDFVHRDLIEVITR 737
 KH L Q G + K F +LLGT FWEGVDF + +I R
 Sbjct: 775 KHNMSASGIQLLAQGIGTGGSPGKLMKTFKTSNQAILLGTNHFWEVDFPGDELTTVMIVR 834

Query: 738 LPFDTPKDYFIQKLSQSLTKEGKNFFDYSLPMTVLKQLKQALGRITRREEQKSAVILDS 797
 LPF +P + K+GKN F SLP VL +Q +GR R K +IILD
 Sbjct: 835 LPFRSPDHPHAAKCELRKKGKNPFQTVSLPEAVLTFRQIGRLLRSAGDKGTIIILDR 894

Query: 798 RLVIKSYGQTIMHSL 812
 R+ YG+ + +L
 Sbjct: 895 RIKTAGYGRFLDAL 909

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5515> which encodes the amino acid sequence <SEQ ID 5516>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3735(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 500/835 (59%), Positives = 626/835 (74%), Gaps = 2/835 (0%)

Query: 1 MFCEFDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIIDSYETDV 60
 MFCEFDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIIDSYETDV
 Sbjct: 1 MFCEFDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIIDSYETDV 60

Query: 61 NPHESLDEHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAEQLF 120
 NPHESLDEHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAE LF
 Sbjct: 61 NPHESLDEHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAEALF 120

Query: 121 LEGCELRTPRIDTVLSQVFPCELEKYSLGALAESLNIELTDAHTAIADARATAQLFIKL 180
 LEG EL PR+DTVEL+Q+F+P EKY+L L+ LNI+L +AHTAIADARATA LF++L
 Sbjct: 121 LEGYELTIPRVDTVLAQLFFPRFEKYNLSHLSRQLNIDLAEHTAIADARATAILFLRL 180

Query: 181 KAKISSLPKEVLETILTADNLLFESYLLIEEAYQEAQFVNPKEYYFWQGLVLKKEKAVG 240
 KI SLP E LE++L ++D+LLFE+ ++I+E +A +P +Y + ++L K
 Sbjct: 181 LQKIESLPICLESLLVYSDSLFETAMVIOGLAKAKPYDPNKYIKIRQILLPKGSKAL 240

Query: 241 KPKKLSDDFQVNMALLGMDARPKQVVFADLVKAHFNQDTTTFLEAQPLGKTYGYLLPLL 300
 KP ++S F +NMALLG++ RPKQ FA L+ ++ +F+EAQ G+GKTYGYLLPLL
 Sbjct: 241 KPYQISKSFPINMALLGLEERPKQTQFAQLIDEDYHQGVASFIEAQTGIGKTYGYLLPLL 300

Query: 301 DQSQKQQIIVSVPTKILQDQIMAKEIKHIQELFHIPCHSIKGPGRNYLKLDAFYKSLQVQD 360
 + + QIIVSVPTK+LQDQ+MA E+ IQE FHI CHS+KGP NYLKLDF SL D

-2003-

Sbjct: 301 AKEDQNQIIVSVPTKLLQDQLMAGEVAAIQEQFHIACHSLKGPANYLKLDSDADSLDQND 360

Query: 361 RNRLINRFKMQLLVWLTTTGGDLDEIKQKQRLSYFDQLKHDGEVTOSSLFYDLDFWKR 420
 +NRL+NR+KMQLLVWL ET TGDLDDEIKQKQ +YF+QLKHDG++ QSS FYD DFW+

5 Sbjct: 361 QNRLVNRYKMQLLVWLLETKTGDLDEIKQKQRFAYFEQLKHDGDIKQSSSEFYDYDFWRV 420

Query: 421 SYDKVAQSQSLVIINHAYFLERVQDDKDFAKGKVLVFDEAQLVLGLNFSRGQLDISHQL 480
 SY+K ++L+I NHAYFL RVQDDKDF+ KVLVFDEAQL+L L+ SR QL+++ L

10 Sbjct: 421 SYEKAKTARLLITNHAYFLHRVQDDKDFARNKVLVFDEAQLMLQLDQLSRHQLNLTVFL 480

Query: 481 QVIQKIIDSSIPLLQKRLLSESISYELSHAVELFYRHNSFEFSETWLKRLKNSINALEVVG 540
 Q IQ + + +PIL+KRILLES+S+EL +Y++ + + W R+ L

Sbjct: 481 QTIQAKLSNPLPLEKRILLESLSFELGQVSSDYQYQNEHQLAHDW-SRIAGYAKELTGAD 539

15 Query: 541 LDELQTFFTATYTYNYWFETDKVNEKRLTILRGAREDFLKFSKFLPPTKTYMISATLQIS 600
 ELQ FF + +YW ++K EKR+T L A + F+ F + LP T KTY +SATL IS

Sbjct: 540 YQELQAFATSDGDYWLSSSEKQEEKRVTYLNSASKAFIHQQLLPETVKTYFVSATLITIS 599

20 Query: 601 PKVYLSDLLGGFSSISTEKIAHEKNANQKVWIDTSMFNILDLSPQYAYEIAKRLQDINT 660
 +V L+DLL GF I +K +Q V +D P + ++S + Y IAKR++ +

Sbjct: 600 SEVTILADLL-GFEEYLYHVIEKDKKQDQLVLVDQEAPIVTEVSDQIYVEAIAKRIESLKQ 658

Query: 661 LKQPTLVLLTSKQTMFMVSDYLDKWEIKHLTQDKNGLAYNVKKRFRGESNLLGTGSFW 720
 P LVL SK+ + +VSDYLD+W++ HL Q+KNG AYN+KKRFD+GE +LLG GSF

25 Sbjct: 659 EGYPIVLVFNKKHLLLVSDYLDQWQVPHLAQEKNGTAYNIKKRFDQGEQTILLGLGSFW 718

Query: 721 EGVDFVHRDRLEIVITRLPFDTPKDYFIQKLSQSLTKEGKNFFDYSLPMTVLKQALG 780
 EGVDF+ DR+I +I RLPFD P+D+F++K+S L ++GKN F DY LPMT+L+LKQA+G

30 Sbjct: 719 EGVDFIQADRMITLIARLPFDNPEDFFVKKMSHYLLEKGNPFRDYFLPMTILRLKQAIG 778

Query: 781 RTTRREEQKSAVILDSRLVIKSYGQTIMHSLGRDFEISKEKINKVLTEMAKFLI 835
 RT RR++QKS VIILDR RL+ KSYGQ I+ LG++F IS++ + L E FLI

Sbjct: 779 RTMRQDQKSVIILDRRLTKSYGQVILEGLGQEFLLISQQNFHDCLVETDCFLI 833

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1774

A DNA sequence (GBSx1881) was identified in *S.agalactiae* <SEQ ID 5517> which encodes the amino acid sequence <SEQ ID 5518>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9633> which encodes amino acid sequence <SEQ ID 9634> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF12702 GB:AF035157 aspartate aminotransferase [Lactococcus
 lactis]
 Identities = 270/391 (69%), Positives = 314/391 (80%)

Query: 7 MTYLSERVLMMEESVTLAAGAKARELRVQGRDILSLTLGEPDFATPKNIQQAIEAITDG 66
 M S+ VL M+ESVTLAA +A+ L+ QGRDI+ LTLG+PDF TPK I QAAIEAI +G

Sbjct: 1 MKKCSDFVLKMDSEVTLAAANRAKALKAQGRDIIDLTLGQPDFPTPKKIGQAIEAINN 60

Query: 67 RASFYTPSSGLPELKSAINAYFERFYGYSLKPNQVVVGTAKFILYTFMTVLNPGDEVI 126
 +ASFYT + GLPELK A+ Y+ RFY Y ++ N++++ GAKF LY +FM ++P DEVI

-2004-

Sbjct: 61 QASFYTQAGGLPELKKAVQHYWTRFYAYEIQTNEILITAGAKFALYAYFMATVDPLDEVI 120

Query: 127 IPTPYWVSADQIKMAEGKPVFVTAKEVNHFKVTVEQLEAVRTDKTKVILLNSPSNPTGM 186
IP PYWVSY DQ+KMA G PV V AK+ N+FKVTVEQLE RT KTK++LLNSPSNPTGM

5 Sbjct: 121 IPAPYWVSVDQVKMAGGNPVIVEAKQENNFKVTVEQLEKARTSKTKILLNSPSNPTGM 180

Query: 187 IYKAELEAIGNWAVEHDILILADDIYGRLVYNGNIFTPISSLSESIRNQTIVINGVSKT 246
IY EEL AIG WAV HD+LILADDIY RLVYNG FT ISSLS+ IRN+T VINGVSKT

10 Sbjct: 181 IYSKEELTAIGEWAVAHDLILADDIYHRLVYNGAEFTAISLSDEIRNRTTVINGVSKT 240

Query: 247 YAMTGWRVGFVGNHDIIAAMSKVVSQTTSNLTAVSQYATIEALNGSQESFEKMRLAFEE 306
+AMTGWR+G AVG+ +IIAAM+K+ SQTTSN TAV+QYA IEA + +SFEKM AFEE

Sbjct: 241 FAMTGWRIGLAVGDPEIIAAMTKIASQTTSNPTAVAQYAAIEAFEENDKSFEKMHAAFE 300

15 Query: 307 RLNIYPLLQVPGFEVVKPQGAFYLFNPVTKAMEMKGYTDVTAFTDAILEEVGLALVTG 366
RLN IY L +VPGFE+VKP GAFYLF P VTKAM MKGYTDVT FT AILEE G+ALVTG

Sbjct: 301 RLNKIYLQLSEVPGFELVKPNGAFYLFKVTKAMAMKGYTDVTDFTTAILEEAGVALVTG 360

Query: 367 AGFGAPENVRLSYATDLETLKEAVRRLHVFM 397
AGFG+PENVRLSYAT LETL+ AV RL +M

20 Sbjct: 361 AGFGSPENVRLSYATSLETLEAAVTRLDWM 391

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1005> which encodes the amino acid sequence <SEQ ID 1006>. Analysis of this protein sequence reveals the following:

25 Possible site: 30
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.48 Transmembrane 95 - 111 (95 - 113)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 301/397 (75%), Positives = 343/397 (85%)

Query: 7 MTYLSERVLNMEESVTLAAGAKARELRVQGRDILSLTLGEPDFATPKNIQQAAIEAITDG 66
M LS+RVL M+ESVTLAAGA+A+ L+ QGRD+L+LTLGEPDF TPK+IQ AIE+I +G

40 Sbjct: 1 MPKLSKRVLMEKESVTLAAGARAKALKAQGRDVLNLTGEPDFFTPKHIQDKAIESIQNG 60

Query: 67 RASFYTPSSGLPELKSAINAYFERFYGYSLKPNQVVVGTA GAKFILYTFMTVLNPGDEVI 126
ASFYT +SGLPELK+AI Y + YGY L P+Q+V GTGAKFILY FFM VLNPGD+V+

Sbjct: 61 TASFYTNASGLPELKAATYLNQYGYHLSPDQIVAGTGAKFILYAFFMAVLNPGDQVL 120

45 Query: 127 IPTPYWVSADQIKMAEGKPVFVTAKEVNHFKVTVEQLEAVRTDKTKVILLNSPSNPTGM 186
IPTPYWVSY+DQ+KMAEG+P+PV E N FKVTV+QLE RT KTKV+L+NSPSNPTGM

Sbjct: 121 IPTPYWVSYSQVQKMAEGQPIFVQGLEENQFKVTVDQLERARTSKTKVILLNSPSNPTGM 180

50 Query: 187 IYKAELEAIGNWAVEHDILILADDIYGRLVYNGNIFTPISSLSESIRNQTIVINGVSKT 246
IY AEEL AIG WAV +DILILADDIY L VYNGN F PIS+LSE+IR QTI +NGV+K+

Sbjct: 181 IYGAELRAIGEWAVHNDILILADDIYGLVYNGNQFVPISTLSEAIRRQTITVNGVAKS 240

Query: 247 YAMTGWRVGFVGNHDIIAAMSKVVSQTTSNLTAVSQYATIEALNGSQESFEKMRLAFEE 306
YAMTGWRVGF G +II+AMSK++ QTTSNLT VSQYA IEA GSQ S E+MRLAFEE

55 Sbjct: 241 YAMTGWRVGF AAGEPEIISAMSKIIGQTTSNLTTVSQYAAIEAFCSQSLEEMRLAFEE 300

Query: 307 RLNIYPLLQVPGFEVVKPQGAFYLFNPVTKAMEMKGYTDVTAFTDAILEEVGLALVTG 366
RLNI YPLLQVPGFEVVKPQGAFY FPNV KAMEM G++DVT+F +AILEEVGLA+V+G

60 Sbjct: 301 RLNITYPLLQVPGFEVVKPQGAFYFFPNVKKAMEMTGFSVDVTSFANAILEEVGLAVVSG 360

Query: 367 AGFGAPENVRLSYATDLETLKEAVRRLHVFMGSNEIN 403
AGFGAPENVRLSYATD+ETLKEAVRRLHVFM SNEIN

Sbjct: 361 AGFGAPENVRLSYATDIETLKEAVRRLHVFMESNEIN 397

-2005-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1775

A DNA sequence (GBSx1882) was identified in *S.agalactiae* <SEQ ID 5519> which encodes the amino acid sequence <SEQ ID 5520>. This protein is predicted to be asparaginyl-tRNA synthetase (asnS). Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1488(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05415 GB:AP001512 asparaginyl-tRNA synthetase [Bacillus halodurans]
Identities = 252/442 (57%), Positives = 316/442 (71%), Gaps = 15/442 (3%)

Query: 7 SIVDVKDYVQEVETIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYGEESGLE 66
+I + YV QEVT+GAW+ANK GKIAF+QLRDG+ F QGV K E G E
Sbjct: 4 TIAKIGQVVDQEVTLGAWLANKRSSGKIAFLQLRDGTGFIQGVVKA-----EVGDE 55

Query: 67 KFDVIKRLNQETSYYVTGIVKEDERSKFGYELDITDLEVIGESHEYPTPKHEGTDFLMD 126
F K L QE+S+YVTGIV++DER+ GYEL +T ++I E+ +YPITPKEHGT+FLMD
Sbjct: 56 WFQKAKNLTQESSLYVTGIVRKDERAPSGYELTVTSFDIHEATDYPITPKEHGTEFLMD 115

Query: 127 NRHLWLRSRKQMAVMQIRNAIIYSTYEFFDQNGFIKFDSPILSENAEDSTELFETDYFG 186
+RHLW+RSRKQ AV++IRN II +TYEFF +NGF+K D PIL+ +A E +TELF T YF
Sbjct: 116 HRHLWIRSRKHAVLRIRNEIIRATYEFFHENG FVKVDPPIITGSAPEGTTELFHTKYFD 175

Query: 187 KPAFLSQSGQLYLEAGAMALGRVDFGPFVFAEKSKTRRHLTEFWMMDAEYSFLSHEESL 246
+ AFLSQSGQLY+EA A+A GRVF FGP FRAEKSKTRRHL EFWM++ E +F+ EESL
Sbjct: 176 EDAPLSQSGQLYMEAAALAFGRVFSFGPTFRAEKSKTRRHLIEFWMIPEMAFVEFEESL 235

Query: 247 DLQEAYVKALIQGVLDRAPOALDILERDVEALKRYIAEPFKRVSYYDAITLLQEHEADED 306
++QE YV ++Q VL L L RD L+ I PF R+SYDDAI L E D+
Sbjct: 236 EIQENYVAYIVQSVLKHCAIELKTLGRDTSVLES-IQAPFPRI SYDDAIKFLHEKGFDD- 293

Query: 307 TDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKA FYMKPVGNPVERVLCADLLAP 366
+E GDDFG+PHET I+ +F P F+ +YP S K FYM+P P + VLCADL+AP
Sbjct: 294 -----IEWGDDFGAPHETAIAEHFDKPVFITHYPTSLKPFYMEPDNRRDDVVL CADLIAP 348

Query: 367 EGYGEIIGGSMREDDYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGIERMVT 426
EGYGEIIGGS R DYD L +++E + Y +YLDLRKYGSVPH GFG+G+ER V +
Sbjct: 349 EGYGEIIGGSQRISDYDLLKKRLEHDL SLDAYAWYLDLRKYGSVPHSGFGLGLERTV 408

Query: 427 VAGTKHIREAIPFPRMLHRIKP 448
++G H+RE IPFPR+L+R+ P
Sbjct: 409 ISGAGHVRETIPFPRLLNRLYP 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5521> which encodes the amino acid sequence <SEQ ID 5522>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1488(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2006-

An alignment of the GAS and GBS proteins is shown below.

Identities = 443/448 (98%), Positives = 447/448 (98%)

```

5  Query: 1  MSKKLISIVDVKDYVGQEVTTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYG 60
    Sbjct: 1  MSKKLISIVDVKDYVGQEVTTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYG 60

10 Query: 61  EESGLEKFDVIKRLNQETSVYVTGIVKEDERSKFGYELDITDLEVIGESHEYPTPKKEHG 120
    Sbjct: 61  EESGLEKFDVIKRLNQETSVYVTGIVKEDERSKFGYELDITDLEIIGESHEYPTPKKEHG 120

15 Query: 121 TDFLMDNRHLWLRSRKQMAVMQIRNAIIYSTYEFFDQNGFIKFDSPILSENAEDSTELF 180
    Sbjct: 121 TDFLMDNRHLWLRSRKQMAVMQIRNAIIYATYEFFDQNGFIKFDSPILSENAEDSTELF 180

20 Query: 181 ETDYFGKPAFLSQSGQLYLEAGAMALGRVDFGPFVRAEKSSTRRLTEFWMMDAEYSFL 240
    Sbjct: 181 ETDYFGKPAFLSQSGQLYLEAGAMALGRVDFGPFVRAEKSSTRRLTEFWMMDAEYSFL 240

25 Query: 241 SHEESLDLQEAYVKALIQGVLDRAPOALDILERDVEALKRYIAEPFKRVSYDDAITLLQE 300
    Sbjct: 241 SHEESLDLQEAYVKALIQGVLDRAPOALDILERDVEALKRYITEPFKRVSYDDAITLLQE 300

30 Query: 301 HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVFNYPASFKA FYMKPVPGNPERVLC 360
    Sbjct: 301 HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVFNYPASFKA FYMKPVPGNPERVLC 360

35 Query: 361 ADLLAPEGYGEIIGGSMREDDYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI 420
    Sbjct: 361 ADLLAPEGYGEIIGGSMREDNYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI 420

40 Query: 421 ERMVTFVAGTKHIREAIPFPRMLHRIKP 448
    Sbjct: 421 ERMVTFVAGTKHIREAIPFPRMLHRI+P
    Sbjct: 421 ERMVTFVAGTKHIREAIPFPRMLHRIRP 448

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1776

A DNA sequence (GBSx1883) was identified in *S. agalactiae* <SEQ ID 5523> which encodes the amino acid sequence <SEQ ID 5524>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 17
    >>> Seems to have a cleavable N-term signal seq.
        INTEGRAL    Likelihood = -6.85    Transmembrane    103 - 119 ( 102 - 127)
        INTEGRAL    Likelihood = -5.04    Transmembrane    73 - 89 ( 68 - 93)
45    INTEGRAL    Likelihood = -4.19    Transmembrane    31 - 47 ( 31 - 49)
        INTEGRAL    Likelihood = -1.86    Transmembrane    157 - 173 ( 157 - 173)

    ----- Final Results -----
50    bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

55 >GP: AAD40355 GB: AF036485 hypothetical protein [Plasmid pNZ4000]
    Identities = 39/135 (28%), Positives = 72/135 (52%), Gaps = 4/135 (2%)

    Query: 3  KSPARLISFISIAIAINLVGANLALFLRLPTIYLDITIGTLIIAIVILGPWYAASTAFLSALI 62
        K A ++ I A+ IN V LA L+LP++L ++GT L +++ GP A + F++ +I
    Sbjct: 15  KLSAATMTLIPAAVGINIVAKALAEGLKLPVWLGSGLTFLASMLAGPVAGAISGFINNVI 74

60    Query: 63  NWMTTDIFSLYSPVAIVVAIITGILIKRNCKPSS--LLWKSLIISLPGTIIASVITVIL 120

```

-2007-

```

      +T   S  Y+  +I + I  G+L      S+  +   ++II++  +I++ + VI
Sbjct: 75  YGLTLSPISTVYAITSIGIGIAGVGLHANGWFSSARRVFSALIIIAIVSAVISTPLNVIF 134

Query: 121  FKGIT--SSGSSIIA 133
      + G T  + G S+ A
Sbjct: 135  WGGQTGIAWGDLSLFA 149

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1777

A DNA sequence (GBSx1884) was identified in *S.agalactiae* <SEQ ID 5525> which encodes the amino acid sequence <SEQ ID 5526>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1873(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC75223 GB:AE000305 orf, hypothetical protein [Escherichia coli K12]
Identities = 97/305 (31%), Positives = 160/305 (51%), Gaps = 10/305 (3%)

Query: 1  MNKEKIIIDCDPGIDDTLALMYAIQHPKLEVVAITITAGNSPVELGLKNTFVTLLELLNRH 60
      M K KII+DCDEPG DD +A+M A +HP +++++ ITI AGN ++ L N + L
Sbjct: 1  MEKRKIILDCDPGHDDAIAIMMAAKHPAIDLGLGITIVAGNQTLDKTLINGLNVCCQL-EI 59

Query: 61  DIPVYVGDNLPLQREFVSAQDTHGMDGLGENNFTLAQPIIFQEEASADC---FLANYFEHK 117
      ++PVY G P+ R+ + A + HG GL F +P+ Q ES +
Sbjct: 60  NVPVYAGMPQPIMRQQIVADNIHGETGLDGPVF---EPLTRQAESTHAVKYIIDTLMASD 116

Query: 118 NDTSIIALGPLTNIAARALQTNPKLGKHKRFISMGGSFKSHGNCSFVAEYNYWCDPHAAQ 177
      D +++ +GPL+NIA A++ P + + + MGG++ + GN +P AE+N + DP AA+
Sbjct: 117 GDITLVPVGPLSNIAVAMRQPAILPKIREIVLMGGAYGT-GNFTPSAEFNI FADPEAAR 175

Query: 178 YVFENLDKKIEMVGLDITRHLVTPNHL SYMERINPDVSSFIQKITKIFYFDFHWQYEHII 237
      VF + + M+GLD+T V TP+ ++ MER I F ++ +
Sbjct: 176 VVFTS-GVPLVMMGLDLTNQTCTPDVIARMERAGGPAGELFSDIMNFTLKTQFENYGLA 234

Query: 238 GCVINDPLAIAFYVNIENIATGFDSYTDVACH-GIAMGQITVDQYHFYKDKANSKILTSVN 296
      G ++D I Y +N + + Y +V + G G+T+ D+ K AN+K+ +++
Sbjct: 235 GGPVHDATCIGYLINPDGIKTQEMYVEVDVNSGPCYGRVTCDELGVLGKPANTKVGITID 294

Query: 297 TNLEW 301
      T+ FW
Sbjct: 295 TDWFW 299

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1778

A DNA sequence (GBSx1885) was identified in *S.agalactiae* <SEQ ID 5527> which encodes the amino acid sequence <SEQ ID 5528>. Analysis of this protein sequence reveals the following:

-2008-

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1860(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB62728 GB:AL133423 hypothetical protein SC4A7.24c
 [Streptomyces coelicolor A3(2)]
 Identities = 36/134 (26%), Positives = 57/134 (41%), Gaps = 7/134 (5%)

15 Query: 1 MLYEVTSSNTQGVDGKVYLSNGKIVETNHPLNHL----PGFNPEELIALAWSTCLNATIK 56
 +LY ++ G DG+V +G++ +P + G NPE+L A +S C +
 Sbjct: 8 VLYTAVATAENGRDGRVATDDGRLDVVVNPPKEMGGNGAGTNPEQLFAAGYSACFQGALG 67

20 Query: 57 AILEQKGFKDLKSRVDVTCQLMKEKQVGKGFYFQVNAVASIEKLSLSDSKLIVNKAHSRC 116
 + Q+G S V + K GF V A I + + ++ +V KAH C
 Sbjct: 68 VVARQEGADISGSTVTAKVGIGKND---GFGIIVEISAEIPTVDAATARSLVEKAHQVC 124

Query: 117 PISKLISNAKTINL 130
 P SK T+ L
 Sbjct: 125 PYSKATRGNITVTIL 138

25

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1779

30 A DNA sequence (GBSx1886) was identified in *S.agalactiae* <SEQ ID 5529> which encodes the amino acid sequence <SEQ ID 5530>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.0531(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9635> which encodes amino acid sequence <SEQ ID 9636> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB15482 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 164/285 (57%), Positives = 207/285 (72%), Gaps = 2/285 (0%)

Query: 6 IKLVIVTGMMSGAGKTVAIQSFEDLGYFTIDNMPPTLVPKFLELAAQSGDT-SKIAMVVDM 64
 I+LVI+TGMSGAGKTVAIQSFEDLGYF +DN+PP+L+PKFLEL +S SK+A+V+D+
 Sbjct: 9 IQLVITGMMSGAGKTVAIQSFEDLGYFCVDNLPPSLLPKFLELMKESNSKMSKVALVMDL 68

50 Query: 65 RSRFFFREINSILDSLEINDNINFKILFLDATDTLVSRYSKETRRSHPLAADGRVLDGIS 124
 R R FF + LD + N I +ILFLDA D+ LV+RYKETRRSHPLAA G L+GI+
 Sbjct: 69 RGREFFDRLEALDEMAENPWITPRILFLDAKDSILVTRYKETRRSHPLAATGLPLEGIA 128

55 Query: 125 LERELLAPLKSMSONVVDTSSELTPRQLRKVISKEFSNQDSQSSFRIEVMFSGFKYGIPLD 184
 LERELL LK SQ + DTS++ PR LR+ I K F+ ++ F + VMSFGFKYGIPI+D
 Sbjct: 129 LERELLEELKGRSQIYDTSMDKPRDLREKIVKHFATNQGET-FITVNVMSFGFKYGIPIID 187

Query: 185 ADLVFDVRFLENPYKPELRDKTGLDTEVDYVMSFDESDDFYDHLALIKPILEPGYQNE 244

-2009-

ADLVFDVRFLPNPYY +R TG D EV YVM ++E+ F + L+ L+ +LP Y+ E
 Sbjct: 188 ADLVFDVRFLPNPYYIESMRPLTGKDKEVSSYVMKWNETQKFNEKLIDLLSFMLPSYKRE 247

Query: 245 GKSVLTVAIGCTGGQHRSTAFARHLSKADWTVNESHDRDNKR 289
 GKS + +AIGCTGGQHRS A L++ K D+ + +HRD KR
 Sbjct: 248 GKSQVVAIGCTGGQHRSVTLAENLADYFKKDYTHVTHRDIEKR 292

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5531> which encodes the amino acid sequence <SEQ ID 5532>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15482 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 164/291 (56%), Positives = 213/291 (72%), Gaps = 3/291 (1%)

Query: 1 MSDKH-INLVIVTGMMSGAGKTVAIQSFEDLGFTIDNMPPALVPKFLELIEQTNENR-RV 58
 +S+ H I LVI+TGMMSGAGKTVAIQSFEDLGTF +DN+PP+L+PKFLEL++++N +V
 Sbjct: 3 VSESHDIQLVITGMMSGAGKTVAIQSFEDLGTFVCDNLPPSLPKFLELMKESNSKMSKV 62

Query: 59 ALVDMRSRLFFKEINSTLDSIESNPSIDFRILFLDATDGLVSRKETRSHPLAADGR 118
 ALV+D+R R FF + LD + NP I RILFLDA D LV+RYKETRRSHPLAA G
 Sbjct: 63 ALVMDLRGREFFDRLEALDEMAENPWITPRILFLDAKDSILVTRYKETRRSHPLAATGL 122

Query: 119 VLDGIRLERELLSPLKSMQHVVDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMVSFGFK 178
 L+GI LERELL LK SQ + DT+ + PR LR+ I F+ + +F + VMSFGFK
 Sbjct: 123 PLEGIALERELLEELKGRSQIYYDTSMDKPRDLREKIVKHFATNQGE-TFTVNMVSFGFK 181

Query: 179 YGLPLDADLVFDVRFLPNPYYQVELREKTGLDEDVFNVMSEVSEVFKHLLNLIVPIL 238
 YG+P+DADLVFDVRFLPNPYY +R TG D++V +YVM E++ F + L++L+ +L
 Sbjct: 182 YGIPIDADLVFDVRFLPNPYYIESMRPLTGKDKEVSSYVMKWNETQKFNEKLIDLLSFML 241

Query: 239 PAYQKEGKSVLTVAIGCTGGQHRSVAFACLAESLATDWSVNESHDRDNRR 289
 P+Y++EGKS + +AIGCTGGQHRSV A LA+ D+ + +HRD +R
 Sbjct: 242 PSYKREGKSVVIAIGCTGGQHRSVTLAENLADYFKKDYTHVTHRDIEKR 292

An alignment of the GAS and GBS proteins is shown below.

Identities = 234/296 (79%), Positives = 263/296 (88%)

Query: 1 MSDEQIKLVIVTGMMSGAGKTVAIQSFEDLGFTIDNMPPTLVPKFLELAAQSGDTSKIAM 60
 MSD+ I LVIVTGMMSGAGKTVAIQSFEDLGFTIDNMPP LVPKFLEL Q+ + ++A+
 Sbjct: 1 MSDKHINLVIVTGMMSGAGKTVAIQSFEDLGFTIDNMPPALVPKFLELIEQTNENRRVAL 60

Query: 61 VDMRSRLFFREINSILDSLEINDNINFILFLDATDTLVSRKETRSHPLAADGRVL 120
 VDMRSRLFF+EINS LDS+E N +I+F+ILFLDATD ELVSRKETRSHPLAADGRVL
 Sbjct: 61 VDMRSRLFFKEINSTLDSIESNPSIDFRILFLDATDGLVSRKETRSHPLAADGRVL 120

Query: 121 DGISLERELLAPLKSMQNVVDTTSELTPRQLRKVISKEFSNQDSQSSFRIEVMVSFGFKYG 180
 DGI LERELL+PLKSMQ+VVDT++LTPRQLRK IS +FS +Q+SFRIEVMVSFGFKYG
 Sbjct: 121 DGIRLERELLSPLKSMQHVVDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMVSFGFKYG 180

Query: 181 IPLDADLVFDVRFLPNPYYKPELRDKTGLDTEVDYVMSFDESDDFYDHLALIKPILPG 240
 +PLDADLVFDVRFLPNPYY+ ELR+KTGLD +V++YVMS ES+ FY HLL LI PILP
 Sbjct: 181 LPLDADLVFDVRFLPNPYYQVELREKTGLDEDVFNVMSEVSEVFKHLLNLIVPILPA 240

Query: 241 YQNEGKSVLTVAIGCTGGQHRSTAFARHLSKADWTVNESHDRDNKRKETVNRS 296
 YQ EGKSVLTVAIGCTGGQHRS AFAC LAE SLATDWSVNESHDRDNKRKETVNRS
 Sbjct: 241 YQKEGKSVLTVAIGCTGGQHRSVAFACLAESLATDWSVNESHDRDNKRKETVNRS 296

-2010-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1780

A DNA sequence (GBSx1887) was identified in *S. galactiae* <SEQ ID 5533> which encodes the amino acid sequence <SEQ ID 5534>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB96620 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 254/321 (79%), Positives = 286/321 (88%), Gaps = 1/321 (0%)

Query: 1 MRKPKITVIGGGTGIPVILKSLRLEDVEITAVVTVADGGSSGELRSVMQ-LTPPGDLRN 59
 MRKPKITVIGGGTGIPVILKSLR +DVEI A+VTVADGGSSGELR MQ LTPPGDLRN
20 Sbjct: 1 MRKPKITVIGGGTGIPVILKSLREKDVEIAAIVTVADGGSSGELRKNMQQLTPPGDLRN 60

Query: 60 VLVALSDMPKFYEQIFQYRFAEGDGFAGHPLGNLIIAGVAEMQGSTYNAMQSLTQFFHT 119
 VLVA+SDMPKFYE++FQYRF+E G FAGHPLGNLIIAG++EMQGSTYNAMQ L++FFHT
25 Sbjct: 61 VLVAMSDMPKFYEVFQYRFSEDAGAFAGHPLGNLIIAGLSEMQGSTYNAMQLLSKFFHT 120

Query: 120 TGKIYPSSEHPLTLHAVFKDGEHVVGESQIADYKGMIDHVYVTNTYNEETPTASRKVVDA 179
 TGKIYPS+HPLTLHAVF+DG EV GES I D++G+ID+VYVTN N++TP ASR+VV
30 Sbjct: 121 TGKIYPSDDHPLTLHAVFQDGEVAGESHIVDHRGIIDNVYVTNALNDDTPLASRRVQT 180

Query: 180 ILESMDIVLGPGLFTSILPNLVIPEIKQALLETRAEVAYVCNIMTQRGETEHFTDADHV 239
 ILESMDIVLGPGLFTSILPN+VI EI +ALLET+AE+AYVCNIMTQRGETEHFTD+DHV
35 Sbjct: 181 ILESMDIVLGPGLFTSILPNIVIKEIGRALLETKAEIAYVCNIMTQRGETEHFTDSHV 240

Query: 240 EVLKRHLGQDAIDTVLVNIEKVPESYMNHFDEYLVQVEHDFSGLRKHARRVISSNFLK 299
 EVL RHLG+ IDTVLVNIEKVP+ YM +N FDEYLVQVEHDF GL K RVISSNFL+
40 Sbjct: 241 EVLHRHLGRPFIDTVLVNIEKVPQEYMNSNRFDEYLVQVEHDFVGLCKQVSRVISSNFLR 300

Query: 300 LEKGGAFHHGDFVVEELMNLV 320
 LE GGAFH GD +V+ELM ++
40 Sbjct: 301 LENGGAFHDGLIVDELMRII 321

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5535> which encodes the amino acid sequence <SEQ ID 5536>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 251/320 (78%), Positives = 284/320 (88%)

55 Query: 1 MRKPKITVIGGGTGIPVILKSLRLEDVEITAVVTVADGGSSGELRSVMQLTPPGDLRNV 60
 M+ PK+TVIGGGTGI +ILKSLR E V+ITAVVTVADGGSSGELR+ MQL PPGDLRNV
 Sbjct: 1 MKNPKMTVIGGGTGISILKSLRNEAVDITAVVTVADGGSSGELRNAMQLAPPGLRNV 60

Query: 61 LVALSDMPKFYEQIFQYRFAEGDGFAGHPLGNLIIAGVAEMQGSTYNAMQSLTQFFHTT 120

-2011-

L+A+SDMPKFYE++FQYRF E DG AGHPLGNLIAG++EMQGSTYNA+Q LT+FFH T
 Sbjct: 61 LLAMSDMPKFYERVFQYRFNESDGLAGHPLGNLIAGISEMQGSTYNAIQILTKFFHIT 120
 Query: 121 GKIYPSSEHPLTLHAVFKDGEVVGESQIADYKGMIDHVYVTNTYNEETPTASRKVVDAI 180
 5 GKIYPSSE LTLHAVFKDGEV GES IA Y GMIDHVYVTNTYN++ P ASRKVV+AI
 Sbjct: 121 GKIYPSSEQALTLHAVFKDGEVAGESSIAKYPGMIDHVYVTNTYNDQKPQASRKVVVEAI 180
 Query: 181 LESDMIVLPGSLFTSILPNLVIPEIKQALLETRAEVAYVCNIMTQRGETEHFTDADHVE 240
 LESDMIVLPGSLFTSILPNLVIPEIK+AL +T+AEV Y+CNIMTQ GETE F+DADHV
 10 Sbjct: 181 LESDMIVLPGSLFTSILPNLVIPEIKALRQTAEVYVICNIMTQYGETEQFSDADHVA 240
 Query: 241 VLKRHLGQDAIDTVLVNIEKVPESYMNHFDEYLVQVEHDFSGLRKHARRVISSNFLKL 300
 VL +HLG+D IDTVLVN+ KVP++YM +N FDEYLVQV+HDF+GL + A+RVISS FL+L
 Sbjct: 241 VLNQHLGRDLIDTVLVNVAKVPQAYMNSNKFDEYLVQVDHDFAGLCRAAKRVISSYFLRL 300
 15 Query: 301 EKGGAFFHGGDFVVEELMNLV 320
 E GGAFH G+ VVEELMNLV
 Sbjct: 301 ENGGAFFHDGNLVVEELMNLV 320

- 20 SEQ ID 5534 (GBS269) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 12; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 5; MW 60.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1781

A DNA sequence (GBSx1888) was identified in *S.agalactiae* <SEQ ID 5537> which encodes the amino acid sequence <SEQ ID 5538>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2479(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB96619 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 209/303 (68%), Positives = 260/303 (84%)

40 Query: 1 MSFTVKVKEELLGHKSENKMSLSAIIKMSGSLGLANHGLNLSITTENAKIARHIYSMLLE 60
 MSFTV VKEE+LG ++ ELSAIIKMSG+GL+ GL LS+ TENAK+ARH+Y
 Sbjct: 1 MSFTVAVKEEILGQHLSRHELSAIIKMSGSIGLSTSGTLTVSVVTENAKLARHLYESFLH 60
 Query: 61 HYHLQPEIKYHQKTNLKRNRYTVTFIEEKVDVILADLKLADAFPGIETGIEHSILDNDEN 120
 45 Y ++ EI++HQ++NLRKNRYTVTF +EKV +L+DL LAD+FFG+ETGI+ +IL ++E
 Sbjct: 61 FYEIKSEIRHHQRSNLRKNRYTVFTDEKVDLLSDHLADSFFGLETGIDEAILSDEEA 120
 Query: 121 GRAYLRGAFLSTGTVPREDSGKYQLEIFSVYLDHAQDLANLKKFMDLAKVIEKHGAVT 180
 GRAYL GAF+ G++R+P+SGKYQLEI SVYLDHAQ +A+L+++F+LDKAV+E K GAVT
 50 Sbjct: 121 GRAYLCGAFLANGSIRDPESGKYQLEISSVYLDHAQGIASLLQQFLLDKAVLERKKGAVT 180
 Query: 181 YLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIARTITASMKTIN 240
 YLQ+AEDIMDFLIVI AM+ARD FE +K++RETRND+NRANN ETANIART++ASMKTIN
 Sbjct: 181 YLQRAEDIMDFLIVIGAMQARDDFERVKILRETRNDLNRANNAETANIARTVSASMKTIN 240
 55 Query: 241 NIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQIADSLETPLSKSGVNHRLRKINKIA 300
 NI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADSL TPL+KSGVNHRLRKINKIA
 Sbjct: 241 NISKIKDIMGLENLPVDLQEVQRLRIQHPDYSIQQLADSLSTPLTKSGVNHRLRKINKIA 300

-2012-

Query: 301 DEL 303
 DEL
 Sbjct: 301 DEL 303

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5539> which encodes the amino acid sequence <SEQ ID 5540>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1698(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 222/303 (73%), Positives = 269/303 (88%)

Query: 1 MSFTVKVKEELLGHKSENKMEISAIKMSGSLGLANHGLNLSITTENAKIARHIYSMLLE 60
 MSFT KVKEEL+ + + EL+AIK+SGSLGLA+ L+LSITTENAKIAR+IYS++E+
 20 Sbjct: 1 MSFTTKVKEELIHLSTGDNNELAAIKLSGSLGLAHQSLHLSITTENAKIARYIYSLIED 60

Query: 61 HYHLOPEIKYHQKTNLKRNRYTVTFIEEKVDVILADLKLADAFFGIETGIEHSILDNDEN 120
 Y + PEI+YHQKTNLKRNRYTV++E+ V+ ILADLKLAD+FFG+ETGIE +L +D
 25 Sbjct: 61 AYVIVPEIRYHQKTNLKRNRYTVYVEQGVTILADLKLADSFFGLETGIEPQVLSDDNA 120

Query: 121 GRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDAKVIEHKGAVT 180
 GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMLDAK IEHK GAVT
 30 Sbjct: 121 GRSYLKGAFLAAGSIRDPESGKYQLEIYSVYLDHAQDLAQLMQKFMLDAKTIEHKGAVT 180

Query: 181 YLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIARTITASMTIN 240
 YLQKAEDIMDFLI+I AM ++ FE IK++RE RNDINRANN ETANIA+TI+ASMTIN
 35 Sbjct: 181 YLQKAEDIMDFLIIGAMSKEDFEAIKLLREARNNDINRANNAETANIAKTISASMTIN 240

Query: 241 NIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQIADSLTPLSKSGVNHRLRKINKIA 300
 NIIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGVNHRLRKINKIA
 35 Sbjct: 241 NIIKIMDTIGLESLPIELQQAQLRVKHPDYSIQQVADALEFPITKSGVNHRLRKINKIA 300

Query: 301 DEL 303
 D+L
 40 Sbjct: 301 DDL 303

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1782

- 45 A DNA sequence (GBSx1889) was identified in *S.agalactiae* <SEQ ID 5541> which encodes the amino acid sequence <SEQ ID 5542>. This protein is predicted to be dipeptidase. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 55 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA86210 GB:Z38063 dipeptidase [Lactobacillus helveticus]
 Identities = 218/473 (46%), Positives = 310/473 (65%), Gaps = 14/473 (2%)

-2013-

Query: 3 CTTILVGKKASYDGMTIARTEDSVNGDFTPKKLVMTSKDQPRHYKSVLSNFEVD---L 59
 CTTILVGKKAS DGMTIAR+ED P+ KV+ +DQP+HY SV+S ++D L
 Sbjct: 6 CTTILVGKKASIDGMTIARSEDG-GRVIIPEGFKVVPEDQPKHYTSVSKQKIDDEDL 64

Query: 60 PDNPLPYTSVPDALGKDGIGWGEAGINSKNVAMSATETITITNSRVLGADPLVSD---GIGE 116
 + PL YTS PD GK+GIWG AGIN+ NVAM+ATETITITNSR+ G DP++ G+GE
 Sbjct: 65 AETPLRYTSAPDVSGKNGIWAAGINADNVAMTATETITITNSRIQGVDPILDPSGGLGE 124

Query: 117 EDILTIVLPYIQSAREGVERLGAILEKYGTYESNGIAFSDTEEIWWLETIGGHHWIARRV 176
 ED +TL LPY+ SA +GV+R+G ++EKYGTYE NG+AFSD + IW+LETIGGHHWIARR+
 Sbjct: 125 EDFVTLTLPYLHSAFDGKRVGYLVEKYGTYESNGMAFSDKDNWILETIGGHHWIARRI 184

Query: 177 PDDVYVITNPQLGIDHFEFNNCDDYMCSSDLKEFIEQYHLDLTYSNEHFNPRIYAFGSQRD 236
 PDD YV PN+L ID F+F++ +++ +SDLK+ I++YHL+ E +N R+ FGS
 Sbjct: 185 PDDAYVIAPNRLNIDTFDDSENFAASDLKDLIDEYHLN--PDREGYNMRHIFGSSTI 242

Query: 237 KDRHYNTPRSWAMQRFNPEIEQDPRSLFIPWCQKPYRKITVEDIKYVLSHDYQDSVYDP 296
 KD HYN PR+W + + +P+ P P+ + R I++EDIK+ S HYQD+ YD
 Sbjct: 243 KDAHYNPNRAWYIHNYPDFDFGGTFADQDQPFICRANRLISIEDIKWAESSHYQDTPYDA 302

Query: 297 YGPEGDAVSRRAFRSVGINRTSQTSLQLRPNKSLETTGVQWLSYGSMPFATMVPLFTQV 356
 YG +G ++ FR +GINR +T ILQ+R + E GVQWL++G F +M+P +T V
 Sbjct: 303 YGDQGTPEQKKTFRPIGINRNFETHILQIRNDVPAEIAQVQWLAFGPNTFNSMLPFYTNV 362

Query: 357 ETVPNYFSNTTKDASTDNFYWTRNLIAALADPHFYQHEADIESYIERTMAQGHADINGVD 416
 T P + T K + + +W N+L A L D ++ + +++ +++AQ H + D
 Sbjct: 363 TTTPEAWQTTPK-FNLNKIFWLNLKTAQLGDTNRYVYGELEDAFEQKSLAQCHKIQHETD 421

Query: 417 REVAENKEIDFQOK----NQEMSDYIQKESQELLNRLIFDASNLMTNRFMSMGD 465
 +EV + Q K NQ+MSD + + ELL +++ + LMT ++ + D
 Sbjct: 422 KEVKNLSGKELQDKLIAANQKMSDTVYNNTVELLGQMVDEGHGLMTLKYDILLD 474

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5543> which encodes the amino acid
 sequence <SEQ ID 5544>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0514(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 345/464 (74%), Positives = 407/464 (87%)

Query: 2 ACTTILVGKKASYDGMTIARTEDSVNGDFTPKKLVMTSKDQPRHYKSVLSNFEVDLPD 61
 +CTTILVGKKASYDGMT+ARTEDS NGDFTPKK+ V+ +DQPRHY+SV S+FE+DLPD
 Sbjct: 9 SCTTILVGKKASYDGMTVARTEDSQNGDFTPKKMIVVKPEDQPRHYRSVQSSFEMDLPD 68

Query: 62 NPLPYTSVPDALGKDGIGWGEAGINSKNVAMSATETITITNSRVLGADPLVSDGIGEEIDLT 121
 NP+ YTSVPDALGKDGIGW EAG+N NVAMSATETITITNSRVLGADPLV+ GIGEEID++T
 Sbjct: 69 NPMTYTSVPDALGKDGIAEAGVNEANVAMSATETITITNSRVLGADPLVASGIGEEIDMVT 128

Query: 122 LVLPLYIQSAREGVERLGAILEKYGTYESNGIAFSDTEEIWWLETIGGHHWIARRVPDDVY 181
 LVLPLYI+SAREGV RLGAILE YGTYESNG+AFSD +IWWLETIGGHHWIARRVPDD Y
 Sbjct: 129 LVLPLYRSAREGVRLGAILEDYGTYESNGVAFSDEHDIWWLETIGGHHWIARRVPDDAY 188

Query: 182 VTNPQNQLGIDHFEFNNCDDYMCSSDLKEFIEQYHLDLTYSNEHFNPRIYAFGSQRDKDRHY 241
 VTNPNQ GIDHFEFNN +DY+CS+DLK+FI+ YHLDLTYS+EHFNPRYAFGSQRDKDR Y
 Sbjct: 189 VTNPNQFGIDHFEFNNPEDYLCADLKDFIDTYHLDLTYSHEHFNPRIYAFGSQRDKDRQY 248

Query: 242 NTPRSWAMQRFNPEIEQDPRSLFIPWCQKPYRKITVEDIKYVLSHDYQDSVYDPYGPPEG 301
 NTPR+W MQ+FLNPEI QDPRS + WCQKPYRKITVED+KYVLS HYQD+ YDPY EG
 Sbjct: 249 NTPRAWIMQKFLNPEIVQDPRSFALAWCQKPYRKITVEDVKYVLSHYQDTGYDPYGSEG 308

-2014-

Query: 302 DAVSRRAFRSVGINRTSQTSLQLRPNKSLLETTGVQWLSYGSMPPFATMVPLFTQVETVPN 361
 VS++ FR +GINRTSQT+IL +RPNK E +QW++YGSMPF TMVP FTQV+T+P+
 Sbjct: 309 TPVSKKVFRPIGINRTSQTALLHIRPNKPQEIAAIQWMAYGSMPFNTMVFFFTQVKTI PD 368

Query: 362 YFSNTTKDASTDNFYWTNRLIAALADPHFYQHEADIESYIERTMAQGHADINGVDREVAE 421
 YF+NT ++ TDNFYWTNRLIAALADPH+ HE D+++Y+E TMA+GHA ++ V+ ++
 Sbjct: 369 YFANTYENVFTDNFYWTNRLIAALADPHYNHETDLDNYLEETMAKGHAMLHAVEVQLLA 428

Query: 422 NKEIDFQQKNQEMSDYIQKESQELNRLIFDASNLMTNRFSGMD 465
 + +D +++NQ+MSDY+Q E+Q LLN+ILFDASNLMTNRFSG+ D
 Sbjct: 429 GETVDLEENQKMSDYVQGETQTLNKLIFDASNLMTNRFSLSD 472

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1783

A DNA sequence (GBSx1890) was identified in *S.agalactiae* <SEQ ID 5545> which encodes the amino acid sequence <SEQ ID 5546>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA96185 GB:Z71552 AdcA protein [Streptococcus pneumoniae]
 Identities = 257/429 (59%), Positives = 312/429 (71%), Gaps = 7/429 (1%)

Query: 1 MRKKFLLLSFVAMFAAWQLVQVKQVWADSKLKVVTTTFYPVYEFTKNVVGDKADVSM LK 60
 M+K LLL S A+F + Q AD KL +VTTTFYPVYEFTK V GD A+V +LI
 Sbjct: 1 MKKISLLASLALFL---VACSNQKQADGKLNIVTTTFYPVYEFTKQVAGDTANVELLIG 57

Query: 61 AGTEPHDFEPSTKNIAAIQDSNAFVYMDNMETWAPKVAKSVKSKVTTIKGTGDM L LTK 120
 AGTEPH++EPS K +A IQD++ FVY ++NMETW PK+ ++ KKV TIK TGDMLL
 Sbjct: 58 AGTEPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDLDKCKVKTIKATGDM L L L P 117

Query: 121 GVEEEGEEHEGHGHEGHHHELDPHVWLSPERAISVVENIRNKFKVAYPKDAASFKNADA 180
 G EEE +H+ HG EGHHE DPHVWLS P RAI +VE+IR+ YP +F KNA A
 Sbjct: 118 GEEEEEGDHD-HGEEGHHHEFDPHVWLSPVRAIKLVEHIRDTLSADYPDKKETFEKNA A 176

Query: 181 YIAKLKELDKKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDAEPSSKRLG 240
 YI KL+ LDK Y GLS AK+KSFVTQHAA F Y+ALDYGL QV I+GL+PDAEPS+ RL
 Sbjct: 177 YIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNALDYGLKQVAISGLSPDAEPSAARLA 236

Query: 241 ELAKYIKKYNINYIYFEENASNKVAKTLADEVGKTA VLSPLEGLSKKEMAAGEDYFSVM 300
 EL +Y+KK I YIYFEENAS +A TL+ E GVKI VL+PLE L++++ AGE+Y SVM
 Sbjct: 237 ELTEYVKNKIAYIYFEENASQALANTLSKEAGVKTDV LNPLESITFEEDTKAGENI SVM 296

Query: 301 RRNLKVLKKT TDVAGKEVAPEE-DKTKTVETGYFKTKDV KDRKLT DYSGNWQSVYPLLQD 359
 +NLK LK+TTD G + PE+ + TKTV+ GYF+ VKDR L+DY+GNWQSVYP L+D
 Sbjct: 297 EKNLQKALQITDQEGPAIEPEKAEDTKTVQNGYFEDA AVKDR T L S D Y A G N W Q S V Y P F L E D 356

Query: 360 GTLDPVWDYKAKSKDMTAAEYKKYYTAGYKTDVESIKIDGKKHQM TFVRNGKSQTFTYK 419
 GT D V+DYKAK MT AEYK YTT GY+TDV I I + M FV+ G+S+ +TYK
 Sbjct: 357 GTFDQVFDYKAKLTGKMTQAEYKAYTYKG YQTDVTKINI--TDNTMEFVQGGQSKKYTYK 414

Query: 420 YAGYKILTY 428
 Y G KILTY
 Sbjct: 415 YVGKKILTY 423

-2015-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5547> which encodes the amino acid sequence <SEQ ID 5548>. Analysis of this protein sequence reveals the following:

Possible site: 17

5 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

10 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA96185 GB:Z71552 AdcA protein [Streptococcus pneumoniae]

Identities = 259/438 (59%), Positives = 326/438 (74%), Gaps = 16/438 (3%)

15 Query: 1 MKKKILLMMSLISVFFAWQLTQAKQVLAEGKVKVVTTFYPVYEFTKGVIGNDGDFVFLMK 60
MKK LL+ SL ++F + + Q A+GK+ +VTFYPVYEFTK V G+ +V +L+
Sbjct: 1 MKKISLLLASLALFL--VACSNQKQADGKLNIVTTFYPVYEFTKQVAGDTANVELLIG 57

20 Query: 61 AGTEPHDFEPSTKDIKKIQDADAFVYMDNMTWVSDVKKSLTSKKVTIVKGTGNMLLVA 120
AGTEPH++EPS K + KIQDAD FVY ++NMTWV + +L KKV +K TG+MLL+
Sbjct: 58 AGTEPHEYPEAKAVAKIQDADTFVYENENMETWVPKLLDTLDDKKVKTIKATGDMMLLLP 117

25 Query: 121 GAGHDHPHEDADKKHEHNKHSEEGHNHAFDPHVWLSFYRSITVVENIRDSLSKAYPEKAE 180
G E+ + H+H EEGH+H FDPHVWLS R+I +VE+IRD+LS YP+K E
Sbjct: 118 GG-----EEEEGDHDHG---EEGHHHEFDPHVWLSFVRAIKLVEHIRDTLSADYDPKKE 168

30 Query: 181 NFKANAATYIEKLKELDKDYTAALSDAKQKSFVTQHAAFGYMALDYGLNQISINGVTPDA 240
F+ NAA YIEKL+ LDK Y LS AK+KSFVTQHAAAF Y+ALDYGL Q++I+G++PDA
Sbjct: 169 TFEKNAAAYIEKLQSLDKAYAEGLSQAQKESFVTQHAAFNALDYGLKQVAISGLSPDA 228

35 Query: 241 EPSAKRIATLSKYVKKYGIKIYIYFEENASSKVAKTLAKEAGVKA AVLSPLEGLTEKEMKA 300
EPSA R+A L++YVKK I YIYFEENAS +A TL+KEAGVK VL+PLE LTE++ KA
Sbjct: 229 EPSAARLAELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDLNPLESLTEEDTKA 288

40 Query: 301 GQDYFTVMRKNETLRLTTDVAGKEILPEK-DTTKT VYNGYFKDKEVKDRQLSDWSGSWQ 359
G++Y +VM KNL+ L+ TTD G I PEK + TKT V NGYF+D VKDR LSD++G+WQ
Sbjct: 289 GENYISVMEKNL KALKQTDDQEGPAIEPEKAEDTKTVQNGYFEDA AVKDR TLDYAGNWQ 348

45 Query: 420 GEKKTFTYTYAGKEILTY 437
G+ K +TY Y GK+ILTY
Sbjct: 406 GQSKKYTYKYVGKKILTY 423

An alignment of the GAS and GBS proteins is shown below.

Identities = 353/515 (68%), Positives = 422/515 (81%), Gaps = 9/515 (1%)

50 Query: 1 MRKKFLLLMMSFVAMFAAWQLVQVQVWADSKLKVVTTTFYPVYEFTKKNVVGDKADVSMLIK 60
M+KK LL+MS +++F AWQL Q KQV A+ K+KVVTTTFYPVYEFTK V+G+ DV ML+K
Sbjct: 1 MKKKILLMMSLISVFFAWQLTQAKQVLAEGKVKVVTTFYPVYEFTKGVIGNDGDFVFLMK 60

55 Query: 61 AGTEPHDFEPSTKNIAAIQDSNAFVYMDNMTWAPKVAKS VKSKVTTIKGTGDMMLLTK 120
AGTEPHDFEPSTK+I IQD++AFVYMDNMTW V KS+ SKKVT +KGTG+MLL
Sbjct: 61 AGTEPHDFEPSTKDIKKIQDADAFVYMDNMTWVSDVKKSLTSKKVTIVKGTGNMLLVA 120

60 Query: 121 GV-----EEEGEEHEGHEGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAA 172
G ++ EH H EGH+H DPHVWLS R+I+VVENIR+ KAYP+ A
Sbjct: 121 GAGHDHPHEDADKKHEHNKHSEEGHNHAFDPHVWLSFYRSITVVENIRDSLSKAYPEKAE 180

Query: 173 SFNKNADAYIAKLKELDKKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDA 232
+F NA YI KLKELDK+Y LS+AKQKSFVTQHAAFGYMALDYGLNQ+ I G+TPDA

-2016-

Sbjct: 181 NFKANAATYIEKLKELDKDYTAALSDAKQKSFVTQHAAFGYMALDYGLNQISINGVTPDA 240

Query: 233 EPSSKRLGELAKYIKKYNINIIYFEENASNKVAKTLADEVGVKTAVLSPLEGLSKKEMAA 292
EPS+KR+ L+KY+KCY I YIYFEENAS+KVAKTLA E GVK AVLSPLEGL++KEM A

5 Sbjct: 241 EPSAKRIATLSKYVKKYGIKIYIYFEENASSKVAKTLAKEAGVKA AVLSPLEGLTEKEMKA 300

Query: 293 GEDYFSVMRRNLKVLKKT'TDVAGKEVAPEEDKTKTVETGYFKTKDKDRKLT'DYSGNWQS 352
G+DYF+VMR+NL+ L+ TTDVAGKE+ PE+D TKT V GYFK K+VKDR+L+D+SG+WQS

10 Sbjct: 301 GQDYFTVMRKNLET'LRLTT'DVAGKEILPEKDT'TKTVYNGYFKDKEVKDRQLSDWSGSWQS 360

Query: 353 VYPLLQDGTLDPVWDYKA-KSKKDMTAAEYKYYTAGYKTDVESIKIDGKKHQM'TFVRNG 411
VYP LQDGTLD VWDYKA KSK MTAAEYK YYT GYKTDVE IKI+GKK MTFVRNG

Sbjct: 361 VYPYLQDGTLDQVWDYKAKKSKGKMTAAEYKDYTTGYKTDVEQIKINGKKKMTMTFVRNG 420

15 Query: 412 KSQTFYTYAGYKILTYKKNRGVRYLFEAKEKDAGQFKYIQFSDHGIKPNKAEHFHIFW 471
+ +TFTY YAG +ILTY KGNRGVR++FEAKE DAG+FKY+QFSDH I P KA+HFH++W

Sbjct: 421 EKQTFYTYAGKEILTYPKGNRGVRFMFEAKEADAGEFKYVQFSDHAIAPKAEKHFHLYW 480

Query: 472 GSESQEKLFEEENWPTYFPAKMSGREVAQDLMSH 506
G +SQEKL +E+E+WPTY+ + +SGRE+AQ++ +H

20 Sbjct: 481 GGDSQEKLHKELEHWPTYYGSDLSGREIAQEINAH 515

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 A related GBS gene <SEQ ID 8899> and protein <SEQ ID 8900> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3
SRCFLG: 0
McG: Length of UR: 19
30 Peak Value of UR: 2.79
Net Charge of CR: 3
McG: Discrim Score: 9.08
GvH: Signal Score (-7.5): 2.59
Possible site: 15
35 >>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 16
ALOM program count: 0 value: 7.69 threshold: 0.0
PERIPHERAL Likelihood = 7.69 264
modified ALOM score: -2.04
40 *** Reasoning Step: 3
Rule gpol
45 ----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the databases:

3758895|emb|CAA96185.1||Z71552 AdcA protein {Streptococcus pneumoniae} >PIR|T46756|T46756
Zn-binding lipoprotein
adcA [imported] - Streptococcus pneumoniae (fragment)

55 Score = 508 bits (1294), Expect = e-143
Identities = 257/429 (59%), Positives = 312/429 (71%), Gaps = 7/429 (1%)

Query: 1 MRKKFLLLLMSFVAMFAAWQLVQVQVWADSKLKVVTTFYPVYEFTKNVVGDKADVSMLIK 60
M+K LLL S A+F + Q AD KL +VTTTFYPVYEFTK V GD A+V +LI

60 Sbjct: 1 MKKISLLLLASLALFL---VACSNQKQADGKLNIVTTTFYPVYEFTKQVAGDTANVELLIG 57

Query: 61 AGTEPHDFEPSTKNIAIQDSNAFVYMDNMETWAPKVAKSVKSKVTTIKGTGDMLLTK 120
AGTEPH++EPS K +A IQD++ FVY ++NMETW PK+ ++ KKV TIK TGDMLL

-2017-

Sbjct: 58 AGTEPEHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKKVKTIKATGDMLLLP 117

Query: 121 GVEEEGEEHGHGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAASFKNADA 180
G EEE +H+ HG EGHHE DPHVWLS RAI +VE+IR+ YP +F KNA A

5 Sbjct: 118 GGEEEEGDHD-HGEEGHHEFDPHVWLSPVRAIKLVEHIRDTLSADYDPKKETFEKNAAA 176

Query: 181 YIAKLKELDKKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDAEPSSKRLG 240
YI KL+ LDK Y GLS AK+KSFVTQHAAF Y+ALDYGL QV I+GL+PDAEPS+ RL

10 Sbjct: 177 YIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDAEPSAARIA 236

Query: 241 ELAKYIKKYNINYIYFEENASNKVAKTLEDEVGVKTAVLSPLEGLSKKEMAAGEDYFSVM 300
EL +Y+KK I YIYFEENAS +A TL+ E GVKT VL+PLE L++++ AGE+Y SVM

15 Sbjct: 237 ELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDLNPLESLTEEDTKAGENYISVM 296

Query: 301 RRNLKVLKKTITDVAGKEVAPEE-DKTKTVETGYFKTKDVKDRKLTDSGNWQSVYPLLQD 359
+NLK LK+ITD G + PE+ + TKT V+ GYF+ VKDR L+DY+GNWQSVYP L+D

20 Sbjct: 297 EKNLKALKQITDQEGPAIEPEKAEDTKTVQNGYFEDA AVKDR T L S D Y A G N W Q S V Y P F L E D 356

Query: 360 GTLDPVWDYKAKSKKDMTAAEYKYYTAGYKTDVESIKIDGKXHQMTFVRNGKSQTFTYK 419
GT D V+DYKAK MT AEYK YYT GY+TDV I I + M FV+ G+S+ +TYK

25 Sbjct: 357 GTFDQVFDYKAKLTGKMTQAEYKAYYTKGYQTDVTKINI--TDNTMEFVQGGQSKKYTYK 414

Query: 420 YAGYKILTY 428
Y G KILTY

Sbjct: 415 YVGKKILTY 423

SEQ ID 8900 (GBS325) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 3; MW 58kDa).

30 The GBS325-His fusion product was purified (Figure 210, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 257A) and FACS (Figure 257B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Example 1784

35 A DNA sequence (GBSx1891) was identified in *S.agalactiae* <SEQ ID 5549> which encodes the amino acid sequence <SEQ ID 5550>. This protein is predicted to be ribosomal protein L31 (rl31). Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1948(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9637> which encodes amino acid sequence <SEQ ID 9638> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF80389 GB:AF160251 ribosomal protein L31 [*Listeria innocua*]
Identities = 61/81 (75%), Positives = 71/81 (87%), Gaps = 1/81 (1%)

50 Query: 9 MKKDIHPDYRPVVFLLDTTGYKFLSGSTKSTKETVEFE-GETYPLIRVEISSDSHPFYTG 67
MK IHP+YRPVVF+DT+T +KFLSGSTKS+ ET+++E G YPL+RVEISSDSHPFYTG
Sbjct: 1 MKTGIHPEYRPVVFVDTSTDFKFLSGSTKSSSETIKWEDGNEYPLLRVEISSDSHPFYTG 60

Query: 68 RQKFTQADGRVDRFNKKYGLK 88
+QK ADGRVDRFNKKYGLK

55 Sbjct: 61 KQKHATADGRVDRFNKKYGLK 81

-2018-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5551> which encodes the amino acid sequence <SEQ ID 5552>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

| | | |
|-------------------------|-------------------------------|---------|
| bacterial cytoplasm --- | Certainty=0.1910(Affirmative) | < succ> |
| bacterial membrane --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial outside --- | Certainty=0.0000(Not Clear) | < succ> |

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/86 (94%), Positives = 86/86 (99%)

Query: 9 MKKDIHPDYRPVVFLDITTTGYKFLSGSTKSTKETVEFEGETYPLIRVEISSDSHPFYTGR 68
M+KDIHPDYRPVVFLDITTTGY+FLSGSTK++KETVEFEGETYPLIRVEISSDSHPFYTGR
Sbjct: 1 MRKDIHPDYRPVVFLDITTTGYQFLSGSTKASKETVEFEGETYPLIRVEISSDSHPFYTGR 60

Query: 69 QKFTQADGRVDRFNKKYGLKDANAAQ 94
QKFTQADGRVDRFNKKYGLKDANAA+
Sbjct: 61 QKFTQADGRVDRFNKKYGLKDANAAK 86

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1785

A DNA sequence (GBSx1892) was identified in *S.agalactiae* <SEQ ID 5553> which encodes the amino acid sequence <SEQ ID 5554>. This protein is predicted to be aspartate aminotransferase (aspC). Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

| | | |
|-------------------------|-------------------------------|---------|
| bacterial cytoplasm --- | Certainty=0.1740(Affirmative) | < succ> |
| bacterial membrane --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial outside --- | Certainty=0.0000(Not Clear) | < succ> |

A related GBS nucleic acid sequence <SEQ ID 9421> which encodes amino acid sequence <SEQ ID 9422> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21948 GB:U32714 aminotransferase [Haemophilus influenzae Rd]
Identities = 200/323 (61%), Positives = 264/323 (80%), Gaps = 1/323 (0%)

Query: 1 MQYYQLQNI-HVDMDDIYVNGVSEGISMSMQALLDNDDEVLPMPDYPLWTACVSLAGG 59
+QYYQ + I ++D+YI NGVSE I+M+MQALL++ DEVLVPMPDYPLWTA V+L+GG
Sbjct: 82 VQYYQSKGILGATVNDVYIGNGVSELITMAMQALLNDGDEVLPMPDYPLWTA AVTLSSG 141

Query: 60 NAVHYICDEEANWYPDIDDIKSKITSKTKAIVLINPNNPTGAVYPREILQEIVDIARQND 119
AVHY+CDE+ANW+P IDDIK+K+ +KTKAIV+INPNNPTGAVY +E+LQEIV+IARQN+
Sbjct: 142 KAVHYLCDEDANWFPTIDDIKAKVNAKTKAIVLINPNNPTGAVYSKELLQEIVETARQNN 201

Query: 120 LIIFSDEYDRLVMDGMEHIPIASIAEDIFTVTLSGLSKSHRICGFRVGMVLSGPRQHV 179
LIIF+DE+YD+++ DG H IA++A D+ TVTL+GLSK++R+ GFR GWM+L+GP+ +
Sbjct: 202 LIIFADEIYDKILYDGA VHHHIAALAPDLLTVTLNGLSKAYRVAGFRQGMILNGPKHNA 261

Query: 180 KGYIEGLNMLANMRLCSNVLAQQVIQTSLGGQSSIDSMLLPGGRIVEQRNYTHKAINIIP 239
KGYIEGL+MLA+MRLC+NV Q IQT+LGG QSI+ +LPGR+ EQRN + I +IP
Sbjct: 262 KGYIEGLDMLASMRICANVPMQHAIQTALGGYQSINEFILPGGRILLEQRNKAYDLITQIP 321

-2019-

Query: 240 GLSAVKPNAGLYLFFPKIDTDMYRIDNDEEFVLNFLKQEKVLLTHGRGFNMNTADHFRIVY 299
 G++ VKP +Y+FPKID + I +DE+ VL+ L+QEKVLL HG+GFN ++ DHFRIV
 Sbjct: 322 GITCVKPMGAMYMPFKIDVKFNIHSDEKMVLDDLRLQEKVLLVHGKGFNWHSPDHFRIVT 381

5 Query: 300 LPRVDELTELQEKMARFLSQYKR 322
 LP V++L E K+ARFLS Y++
 Sbjct: 382 LPYVNQLEEAITKLARFLSDYRQ 404

There is also homology to SEQ ID 3662.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1786

A DNA sequence (GBSx1893) was identified in *S.agalactiae* <SEQ ID 5555> which encodes the amino acid sequence <SEQ ID 5556>. Analysis of this protein sequence reveals the following:

15 Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.02 Transmembrane 164 - 180 (163 - 181)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 10099> which encodes amino acid sequence <SEQ ID 10100> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06181 GB:AP001515 transcriptional pleiotropic repressor
 [Bacillus halodurans]
 Identities = 129/257 (50%), Positives = 181/257 (70%), Gaps = 3/257 (1%)
 30 Query: 23 NLLEKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGNNLLGYAMKYKTN 82
 +LL + RKI +LQ+S + + MA L D+I N +++ G LLG+A+K +
 Sbjct: 2 SLLSRMRKINDMLQKSGVQ---HVNPREMAETLRDVISANIFVVSRRGKLLGFAIKQEIE 58
 35 Query: 83 TDRVEEFFETKQFPDYVKSASRVYDTEANLSVDNDLSIFPVETKENFQDGITTIPIYG 142
 +R+++ E +QFP+ Y +V +T ANL ++++ + FPVE KE F+ G+TTI PI G
 Sbjct: 59 NERMKKMLDRQFPPEEYTTGLFKVEETSANLDINSEFTAFPVENKELFKTGLTTIVPISG 118
 40 Query: 143 GGMRLGTFTIWRNDKEFSDDDLILVEIASTVVGIIQLNLQTNLEENIRKQTAVTMAINT 202
 GG RLGT I+ R + F+DDDLIL E +TVVG+++L+ +T+ +EE R + V MAI++
 Sbjct: 119 GGQRLGTLILARLNSFNDDDLILAEGATVVGMEILHEKTQEIEEARSKAVVQMAISS 178
 Query: 203 LSYSEMKAVALIIGELDGLGRLITASVIADRIITRSVIVNALRKLESAGIIESRSLGMK 262
 LSYSE++AV I ELDG EG L AS IADR+GITRSVIVNALRKLESAG+IESRSLGMK
 45 Sbjct: 179 LSYSELEAVEHIFEELDGKEGLLVASKIADRVGITRSVIVNALRKLESAGVIESRSLGMK 238
 Query: 263 GTYLVKVINNEGIFDKLKE 279
 GTY+KV+N+ +L++
 Sbjct: 239 GTYIKVLNDKFLVELEK 255
 50

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5557> which encodes the amino acid sequence <SEQ ID 5558>. Analysis of this protein sequence reveals the following:

55 Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.64 Transmembrane 144 - 160 (143 - 161)
 ----- Final Results -----

-2020-

bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:CAB13490 GB:Z99112 transcriptional regulator [Bacillus subtilis]
 Identities = 131/255 (51%), Positives = 179/255 (69%), Gaps = 3/255 (1%)

Query: 4 LLEKTRKITSILQRSVDSLETLPYNTMASRLADIIDCNACIINGGGTLLGYAMKYKTNT 63
 10 LL+KTR I S+LQ + + + MA L D+ID N +++ G LLGY++ +
 Sbjct: 3 LLQKTRIINSMLQAAAGK---PVNFKEMAETLRDVIDSNIFVVSRRGKLLGYSINQQIEN 59

Query: 64 DRVEEFFEAKQFPDITYVKAASRVYDTEANLSVENELTIFPVESKDTYPGGLTTIPIYGG 123
 DR+++ E +QFP+ Y K V +T +NL + +E T FPVE++D + GLTTI PI GG
 15 Sbjct: 60 DRMKMLEDRQFPPEYTKNLFNPETSSNLDINSEYTAFFVENRDLFQAGLTTIVPIIGG 119

Query: 124 GMRLGSLIIWRNDEFSDDDLILVEISSTVVGIIQLNLQTNLEDTIRKQTAVNMAINTL 183
 G RLG+LI+ R ++F+DDDLIL E +TVVG+++L + E +E+ R + V MAI++L
 20 Sbjct: 120 GERLGTLLSRLQDQFNDDDLILAEGYATVVGMEILREKAEIEEEARSKAVVQMAISSL 179

Query: 184 SYSEMKAVALGELDGNEGRLTASVIADRIGITRSVIVNALRKLESAGIESRSLGMKG 243
 SYSE++A+ I ELGNEG L AS IADR+GITRSVIVNALRKLESAG+IESRSLGMKG
 Sbjct: 180 SYSELEAIEHIFEELDGNEGLLVASKIADRVGITRSVIVNALRKLESAGVIESRSLGMKG 239

Query: 244 TYLKVINNEGIFAKLK 258
 TY+KV+N +L+
 25 Sbjct: 240 TYIKVLNNKFLIELE 254

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/260 (89%), Positives = 247/260 (94%)

Query: 21 MPNLLKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGGNLLGYAMKYK 80
 MPNLLKTRKITSILQRSVDSL+ ELPYNTMA++LADIIDCNACIINGGG LLGYAMKYK
 35 Sbjct: 1 MPNLLKTRKITSILQRSVDSLETLPYNTMASRLADIIDCNACIINGGGTLLGYAMKYK 60

Query: 81 TNDTRVEEFFETKQFPDYVKSASRVYDTEANLSVDNDLSIFFVETKENFQDGITTIPI 140
 TNDTRVEEFFE KQFPD YVK+ASRVYDTEANLSV+N+L+IFPVE+K+ + G+TTIPI
 Sbjct: 61 TNDTRVEEFFEAKQFPDITYVKAASRVYDTEANLSVENELTIFPVESKDTYPGGLTTIPI 120

Query: 141 YGGGMRLGTFIIWRNDKEFSDDDLILVEIASSTVVGIIQLNLQTNLEENIRKQTAVTMAI 200
 YGGGMRLG+ IIWRND EFSDDDLILVEI+STVVGIIQLNLQTNLE+ IRKQTAV MAI
 40 Sbjct: 121 YGGGMRLGSLIIWRNDEFSDDDLILVEISSTVVGIIQLNLQTNLEDTIRKQTAVNMAI 180

Query: 201 NTLSEYSEMKAVALGELDGLGRLTASVIADRIGITRSVIVNALRKLESAGIESRSLG 260
 NTLSEYSEMKAVALGELDG EGRLTASVIADRIGITRSVIVNALRKLESAGIESRSLG
 45 Sbjct: 181 NTLSEYSEMKAVALGELDGNEGRLTASVIADRIGITRSVIVNALRKLESAGIESRSLG 240

Query: 261 MKGTYLKVINNEGIFDKLKEY 280
 MKGTYLKVINNEGIF KLKE+
 50 Sbjct: 241 MKGTYLKVINNEGIFAKLKEF 260

A related GBS gene <SEQ ID 8901> and protein <SEQ ID 8902> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 55 McG: Discrim Score: -6.84
 GvH: Signal Score (-7.5): -5.37
 Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -2.02 threshold: 0.0
 60 INTEGRAL Likelihood = -2.02 Transmembrane 114 - 130 (113 - 131)
 PERIPHERAL Likelihood = 3.61 179
 modified ALOM score: 0.90

*** Reasoning Step: 3

-2021-

----- Final Results -----

```

bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF02556 (223 - 987 of 1293)

EGAD|13275|BS1617(4 - 255 of 259) cody protein {Bacillus subtilis} OMNI|NT01BS189S cody protein (vegetative protein 286b) (veg286b) GP|535351|gb|AAB03372.1|U13634 CodY {Bacillus subtilis} GP|2633989|emb|CAB13490.1|Z99112 transcriptional regulator {Bacillus subtilis} PIR|S61496|S61496 transcription pleiotropic repressor cody - Bacillus subtilis
%Match = 29.1

```
%Identity = 50.6   %Similarity = 71.5
```

Matches = 128 Mismatches = 71 Conservative Sub.s = 53

177 207 237 267 297 327 357 387
 DCKS*NALI*L*RKTYKG*RKCRILEKTRKITSILQRSVDSLDAELPYNTMAAQLADI IDCNACI INGGGNLLGYAMKY
 |:||| | |:|| : : : || | |:|| | :: | |||::
 MALLQKTRIINSLQAAGK---PVNFKEAAETLRDVIDSNIFVVSSRRGKLLGYINSIQ
 10 20 30 40 50

KTNTDRVEE⁴¹⁷⁴⁴⁷FFETKQFPDY⁴⁷⁷VKSASRVYDTEANLSVDNDL⁵⁰⁷SIFPVETKENFQDGITTIAP⁵³⁷IYGGMRLGTFI⁵⁶⁷IWRNDKEF⁵⁹⁷⁶²⁷

: ||:: :| :|||: | |: | :| :|| :::: : ||| : : || :||| || ||||:|: | :|

QIENDRMKKMLERDFPEEYTKNLFNPVETSSNDINSEYTAFPVENRDLFOAGLTIVPIIGGGERLGTLILSRLODQF

70 80 90 100 110 120 130

657 687 717 747 777 807 837 867
SDDDLILVEIASTVVGIGQLNLTENLEENIRKQTAVTMAINTLSYSEMKAVAAILGELDGLGRLTASVIADRIGITRS
:||||| | :|||::| : ||| | : |||::| ::| : ||| | |||::| |||::|
NDDDLILAEGATVVGMEILREKAEEIEEARSKAVVQMAISSLSYSELAIEHIFIELDGNELLVASKIADRVGITRS

 150 160 170 180 190 200 210

897 927 957 987 1017 1047 1077 1107
VIVNALRKLESAGIIESRSLGMKGYLKVINEGIFDKLKEYN*S*HGTGSSFQFLFWNQEEIRRKMTXXN*LXXLFS*RL
|||||||:|||||||:|:| : : :
VIVNALRKLESAGVIESRSLGMKGYIKVLNNKFLIELENLKSH
230 240 250

SEQ ID 8902 (GBS431) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 7; MW 54kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 6; MW 29kDa).

GBS431-GST was purified as shown in Figure 223, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1787

A DNA sequence (GBSx1894) was identified in *S.agalactiae* <SEQ ID 5559> which encodes the amino acid sequence <SEQ ID 5560>. This protein is predicted to be isochorismatase. Analysis of this protein sequence reveals the following:

Possible site: 35

```
>>> Seems to have no N-terminal signal sequence
```

INTEGRAL Likelihood = -2.81 Transmembrane 126 - 142 (125 - 142)

----- Final Results -----

```

bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2022-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15164 GB:Z99120 similar to pyrazinamidase/nicotinamidase
[Bacillus subtilis]
Identities = 99/181 (54%), Positives = 132/181 (72%)

Query: 1 MTKALISIDYTYDFVADDGKLTAGKPAQSIASAIADVTEKAYRSGDYIFFAIDNHDIGDV 60
M KALI IDYT DFVA DGKLT G+P + I AI ++T++ +GDY+ A+D+HD GD
Sbjct: 1 MKKALICIDYTNDVFASDGKLTCEGEPGRMIEEAIVNLTKEFITNGDYVVLAVDSHDEGDQ 60

Query: 61 FHPESNLFPEHNIKGTSGRNLYGPLGTLYETIKEDSRVFWIDKRHYSAFSGTDLDIRLRE 120
+HPE+ LFP HNIKGT G++LYG L LY+ + + V++++K YSAF+GTDL+++LRE
Sbjct: 61 YHPETRLFPFPHNIKGTGKDLGKLLPLYQKHEHEPNVYMEKTRYSAFAGTDLLELKLRE 120

Query: 121 RRVDTLILTGVLTDICVLHTAIDAYNLGYKIEVPAAAVASLNSNHQWALNHFKTVLGATI 181
R++ L L GV TDICVLHTA+DAYN G++I V AVAS N H WAL+HF +GA +
Sbjct: 121 RQIGELHLAGVCTDICVLHTAIDAYNKGFRIVVHKQAVASFNQEGHAWALSHFANSIGAQV 181

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5561> which encodes the amino acid sequence <SEQ ID 5562>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.60 Transmembrane 126 - 142 (126 - 142)

----- Final Results -----
bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15164 GB:Z99120 similar to pyrazinamidase/nicotinamidase
[Bacillus subtilis]
Identities = 90/179 (50%), Positives = 127/179 (70%)

Query: 3 RALISIDYTNDVFADDGKLSAGKSAQAIATKIAEVTKTAFDQGDYIFFAIDCHDQND SWH 62
+ALI IDYTNDFVA DGKL+ G+ + I I +TK GDY+ A+D HD+ D +H
Sbjct: 3 KALICIDYTNDVFASDGKLTCEGEPGRMIEEAIVNLTKEFITNGDYVVLAVDSHDEGDQYH 62

Query: 63 PESKLF AAHNIKGTGRHLYGPLAEVYSYMKQHPRVFWIDKRYYSASFSGTDLDIRLRERG 122
PE++LF HNIKGT G+ LYG L +Y + P V++++K YSAF+GTDL+++LRER
Sbjct: 63 PETRLFPFPHNIKGTGKDLGKLLPLYQKHEHEPNVYMEKTRYSAFAGTDLLELKLRRERQ 122

Query: 123 ITQLVLTGVLSDICVLHTAIDAYHLGYQLEIVKSAVASLTKESEYEWSLAHFEQVLGAKL 181
I +L L GV +DICVLHTA+DAY+ G+++ + K AVAS +E + W+L+HF +GA++
Sbjct: 123 IGELHLAGVCTDICVLHTAIDAYNKGFRIVVHKQAVASFNQEGHAWALSHFANSIGAQV 181

An alignment of the GAS and GBS proteins is shown below.

Identities = 121/180 (67%), Positives = 150/180 (83%)

Query: 3 KALISIDYTYDFVADDGKLTAGKPAQSIASAIADVTEKAYRSGDYIFFAIDNHDIGDV FH 62
+ALISIDYT DFVADDGKL+AGK AQ+IA+ IA+VI+ A+ GDYIFFAID HD D +H
Sbjct: 3 RALISIDYTNDVFADDGKLSAGKSAQAIATKIAEVTKTAFDQGDYIFFAIDCHDQND SWH 62

Query: 63 PESNLFPEHNIKGTSGRNLYGPLGTLYETIKEDSRVFWIDKRHYSAFSGTDLDIRLRERR 122
PES LF HNIKGT+GR+LYGEL +Y +K+ RVFWIDKR+YSAFSGTDLDIRLRER
Sbjct: 63 PESKLF AAHNIKGTGRHLYGPLAEVYSYMKQHPRVFWIDKRYYSASFSGTDLDIRLRERG 122

Query: 123 VDTLILTGVLTDICVLHTAIDAYNLGYKIEVPAAAVASLNSNHQWALNHFKTVLGATIL 182
+ L+LTGVL+DICVLHTAIDAY+LGY++E+ +AVASL +++W+L HF+ VLGA ++
Sbjct: 123 ITQLVLTGVLSDICVLHTAIDAYHLGYQLEIVKSAVASLTKESEYEWSLAHFEQVLGAKLI 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2023-

Example 1788

A DNA sequence (GBSx1895) was identified in *S.agalactiae* <SEQ ID 5563> which encodes the amino acid sequence <SEQ ID 5564>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1539(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1789

A DNA sequence (GBSx1896) was identified in *S.agalactiae* <SEQ ID 5565> which encodes the amino acid sequence <SEQ ID 5566>. This protein is predicted to be 3-hydroxyacyl-CoA dehydrogenase (hbd-10). Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -0.27 Transmembrane 3 - 19 (1 - 19)

INTEGRAL Likelihood = -0.11 Transmembrane 277 - 293 (277 - 294)

----- Final Results -----

bacterial membrane --- Certainty=0.1107(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF12219 GB:AE001862 3-hydroxyacyl-CoA dehydrogenase, putative

[Deinococcus radiodurans]

Identities = 151/321 (47%), Positives = 196/321 (61%), Gaps = 36/321 (11%)

Query: 56 NMTIKNLTVAGSGVLGSQIAFQAAYKGMVSVTIYDINDEALNKGKERIKKLAKVYQSEIET 115

+M+IK +TV GSGVLGSQIAFQ A+ G V +YDIND A+ K +E + KL YQ +++

Sbjct: 51 SMSIKTVTVCGSGVLGSQIAFQTA FHGFDVHLYDINDAAIAKARETLGKLQARYQQDLKV 110

Query: 116 AKEAYS DKA SIKYKNLPLSLDHIFLSKVADSLDLIADLPNQITFSKNLDQAVSDADLV 175

+ D +I+F ++ +AV DLV

Sbjct: 111 DAQQTGDAFA-----RISFFTDIAEAVKGVDLV 138

Query: 176 IEAVPETVSIKEDFYKQLAKVAPSKTIFATNSSTLVPSQFADITGRDPKFLAMHFANNIW 235

IEA+PE + IK FY QL +VA TIFATNSSTL+PSQF + TGRP+KFLA+HFAN IW

Sbjct: 139 IEAIPENMDIKRFYNQLGEVADPNTIFATNSSTLLPSQFMEETGRPEKFLALHFANEIW 198

Query: 236 QNNIVEIMGHKGTDDVEIKEALAFSKDIGMVPLHIHKEQPGYILNSILVPFLESALALYY 295

+ N EIM TDD V + F+KDIGMV L ++KEQ GYILN++LVP L +AL L

Sbjct: 199 KFNIAEIMRTPRTDDAVFDTVVQFAKDIGMVALPMYKEQAGYILNLTLLVPLLGAALELVV 258

Query: 296 DKVSDSETIDKTWKLGTGAPMGPLEILDIIIGIDTAYNIMKNYSDTNSDPNSLHAHLAKML 355

++D +T+DKTW + TGAP GP LD+IG+ T YNI N + ++P S A AK +

Sbjct: 259 KGIADPQTVDKTMIAATGAPRGPF AFLDVLIGLTPYNI--NMASAE TNP GS--AAAAKYI 314

Query: 356 KEEFIDKGRTGKAAGHGFYDY 376

KE +IDKG+ G A G GFY Y

-2024-

Sbjct: 315 KENYIDKGKLGATATGEGFYKY 335

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8903> and protein <SEQ ID 8904> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
10  McG: Length of UR: 20
      Peak Value of UR: 1.55
      Net Charge of CR: 1
      McG: Discrim Score: -0.60
      GvH: Signal Score (-7.5): -3.93
15  Possible site: 21
    >>> Seems to have no N-terminal signal sequence
    Amino Acid Composition: calculated from 1
    ALOM program count: 1 value: -0.11 threshold: 0.0
      INTEGRAL Likelihood = -0.11 Transmembrane 221 - 237 ( 221 - 238)
20  PERIPHERAL Likelihood = 4.61 6
      modified ALOM score: 0.52
      icml HYPID: 7 CFP: 0.104

*** Reasoning Step: 3
25  ----- Final Results -----
      bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
30  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

37.5/60.5% over 278aa

Archaeoglobus

fulgidus

```
35  EGAD|103851| 3-hydroxyacyl-CoA dehydrogenase Insert characterized OMNI|AF2273 3-
hydroxyacyl-CoA dehydrogenase (hbd-10) Insert
characterized
GP|2648250|gb|AAB88983.1|AE000948 3-hydroxyacyl-CoA dehydrogenase (hbd-10) Insert
40  characterized
PIR|A69534|A69534 3-hydroxyacyl-CoA dehydrogenase (hbd-10) homolog - Insert characterized
```

ORF01176(475 - 1431 of 1731)

```
45  EGAD|103851|AF2273(17 - 295 of 668) 3-hydroxyacyl-CoA dehydrogenase {Archaeoglobus
fulgidus}OMNI|AF2273 3-hydroxyacyl-CoA dehydrogenase (hbd-
10)GP|2648250|gb|AAB88983.1|AE000948 3-hydroxyacyl-CoA dehydrogenase (hbd-10)
{Archaeoglobus fulgidus}PIR|A69534|A69534 3-hydroxyacyl-CoA dehydrogenase (hbd-10) homolog
- Archaeoglobus fulgidus
```

%Match = 14.8

%Identity = 37.5 %Similarity = 60.4

Matches = 106 Mismatches = 106 Conservative Sub.s = 65

```
387      417      447      477      507      537      567      597
KKRYYFKNNHTIYLLLDISFVKLSSKTFNSISIGGCNMTIKNLTVAGSGVLGSQIAFQAAYKGMSTVIYDINDEALNKGK
:      :      || : | |:|:| || | :| |: || | ::|
55  MPRRVKQVINMDVREIRIKTVAVLGAGLMGHGIAEVCAMAGYNVTMRDIKQEFVDRGM
      10      20      30      40      50

624      651      681      711      741      771      801      831
ERIKK-LAKVYQS-EIETAKEAYSDKAKSIKYNKNLLPSLDHIFLSKVADSLDLIADLPNQITFSKNLDQAVSDADLVIE
||: |||: | :|::|:| |
60  NMIKESLAKLEQKGKIKSAEEVLS-----RIKPTVDLEEAVKQDADLVIE
      70      80      90      100
```

861 891 921 951 981 1011 1041 1071
 AVPETVSIKEDFYKQLAKVAPSKTIFATNSSLTLPVSQFADITGRPDKFLAMHFANNIWQNNIVEIMGHKGTDDDEVIKEAL
 |||| | ||: :::: |:| ||:|:|: : :|| | ||:| | :|| | :||:: : | |||: :
 5 AVPEVVEIKKQVWEEVDKLAKPDCITFTSNTSTMRITMLADFTSRPEKFAGLHFFNPPVLMRLVEVIRGEKTSDEVMDDLIV
 120 130 140 150 160 170 180

 1101 1131 1161 1191 1221 1251 1281 1311
 AFSDKIGMVLPHIHKEQPGYILNSILVPFLESALALYDKVSDSETIDKTWKLGTGAPMGPLEILDIIIGIDTAXNIMKNY
 | | || | : : | :|:| : | : : : : | : | | : | | |||:|:| | :| | | : | |
 10 EFVKSIGKTPVRVEKDVPGFIVNRVQAPASVLLMAILEKGIATPPEVDATVR-RLGLPMGPFELVDYTGVDILYNALKYY
 200 210 220 230 240 250 260

 1341 1371 1401 1431 1461 1491 1521 1551
 SDTNSDPNSLHAHLAKMLKEEFIDKGRGTGKAAGHGFYDYD*TIKEVR*KSNLFYNSTKE*LHQBF*NDLKPIDDYHLS
 : | | | : : : || : : : | :| ||| : : : : : : |
 15 AQTIS-PD---YEPKPFLEEMVKANKLGRKTGQGFYDWSKGRPOIDSSKATDKINPMDFTFVEINEAVKLVEMGVATPQ
 270 280 290 300 310 320 330

GBS112-GST was purified as shown in Figure 198, lane 10.

25 A DNA sequence (GBSx1897) was identified in *S.agalactiae* <SEQ ID 5567> which encodes the amino acid sequence <SEQ ID 5568>. Analysis of this protein sequence reveals the following:

```

30      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3332(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

40

Query: 1 MAVLSMLGIIDAKPKVGIFYLQGYHASIGTSHFEKMTVSEIMGIPLTVHQKDSVYDVIVH 60
+A+L+M G ++A+P+VGIFY G+ + +K+ V + IP+ +H+ SVYD I

Sbjct: 43 LAILTMSGFLEARPRVGIFYFTGKTGTOLLADKCLKKLOVKDFOSIPVVIHENVSYDAICT 102

45 Query: 61 IFMEDAGCAFILDDDDFLCGVSRKDLLKISIGGDLKMPIGVMVTRMPHVTTVLENES 120
+F+ED G F++D D L GV+SRKDLL+ SIG +L+ +P+ ++MTRMP++T +
Subject: 103 MFLEDVGTLEFVVDRAVLVGVLRSKDLLRASIGQOELTSVPVHIIMTRMPNITVCRREDY 162

Query: 121 LFAAADKLVSRKVDLPPVRHDKQYPEKFKVIGKLSKLTILASLFLETRD 169
+ A L+ +++D+LPV+ K + F+VIG+++KT + + + + +
Sbjct: 163 VMDIAKHLIEKQIDALPVI---KDTDKGFEVIGRVTKTNMTKILVLSLE 208

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
55 vaccines or diagnostics.

-2026-

Example 1791

A DNA sequence (GBSx1898) was identified in *S.agalactiae* <SEQ ID 5569> which encodes the amino acid sequence <SEQ ID 5570>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.53    Transmembrane    60 - 76 ( 60 - 76)

----- Final Results -----
10      bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:BAB05092 GB:AP001511 unknown conserved protein [Bacillus halodurans]
    Identities = 126/256 (49%), Positives = 183/256 (71%), Gaps = 1/256 (0%)

Query: 7  IFIISDSLGETAKAIAKACLSQFPGHDDWHFQRFYSYINSQERLEQVFEEASQKTVFMMFS 66
      +++SDS+GETA+ + KA SQF G +R Y+ +E +++V + A Q + F+
20  Sbjct: 10 VYVVSDSVGETAELVVKAAASQFSGAGI-EVRRIPYVEDKETVDEVIQLAKQADAIIFT 68

Query: 67 LVDVALASYAQKRCESEHYAYVDLLTNVIQGISRISGIDPLGEPGILRRLDNDYFKRVES 126
      LV + +Y ++ VD++ +++ IS ++ +P EPGI+ RLD DYF++VE+
25  Sbjct: 69 LVPVGIRTYLLEKATEAKVETVDIIGPMLKISSLTKEEPRYEPGIVYRLDEDYFRKVEA 128

Query: 127 IEFVAVKYDDGRDPRGILQADLVIIGISRTSKTPLSMFLADKNIKVINIPLVPEVPVPKEL 186
      IEFVAVKYDDGRDPRGI++ADLV+IG+SRTSKTPLS +LA K +KV N+PLVPEV P+EL
30  Sbjct: 129 IEFVAVKYDDGRDPRGIVRADLVIGVSRTSKTPLSQYLAHKRLKVANVPLVPEVEPPEEL 188

Query: 187 RMDISRRRIIGLTNSVDHLNQVRKVRKSLGLSSTANYASLERILEETRYAEVVMKNLGCP 246
      + +++IGL S + LN +R RLK+LGL S ANYA+++RI EE YAE +MK +GCP
35  Sbjct: 189 FKLSPKKVIGLKISPEQLNGIRAERLKTGLKLSQANYANIDRIKEELAYAE GIMKRIGCP 248

Query: 247 IINVSDKAIEETATII 262
      +I+VS+KA+EETA +I
40  Sbjct: 249 VIDVSNKAVEETANLI 264

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5570 (GBS378) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 2; MW 59kDa).

GBS378-GST was purified as shown in Figure 212, lane 6.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1792

45 A DNA sequence (GBSx1899) was identified in *S.agalactiae* <SEQ ID 5571> which encodes the amino acid sequence <SEQ ID 5572>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3703(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2027-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35361 GB:AE001709 pyruvate,orthophosphate dikinase

[Thermotoga maritima]

Identities = 494/882 (56%), Positives = 639/882 (72%), Gaps = 9/882 (1%)

5

Query: 1 METKFFVYHFD---EGCKEMKELLGGKGANLAEMTSIGLVPVPGGFTITTQACNDYYDNAC 56

M K+VY F EG +MK++LGGKGANLAEMT++G+PVP GFTI+ + C YYD+

Sbjct: 1 MAKKYVYFFANGKAEGRADMKDILGGKGANLAEMTNLGI PVPPGFTISAEVCKYYDHR 60

10

Query: 57 HIRELSIDQAMAQLEVEQNKQLGSVDDPLLVSVRSGSVFSGMMDTVNLGLNDRS 116

E + Q+++AM +LE K+ G ++PLLVSVRSG+ SMPGMDTVNLGLND +

Sbjct: 61 TYPEELKEQVEEAMRRLEEVTGKKFGDPNNPLLVSVRSGAASMPGMDTVNLGLNDET 120

15

Query: 117 VQGLVKKTFEDERFAYDSYRRFIQMFADVVVTGIPKYKFDITLDRKTDKCYQDDTELTS 176

V+GL K T +ERFAYD+YRRF+QMF DVV IP KF+ L+ LK +K + DTEL D

Sbjct: 121 VKGLAKLTNNERFAYDAYRRFLQMFQDVVLKIPHEKFEKALEELKKEGKVLDTLDAED 180

20

Query: 177 LKRLVEFYKELYQKEAGEKFPQDPKRLLLAIEAVFKSWNNPRAKIYRKLNDIPE--TLG 234

LK+LVE YK++Y KE G++FPQDP +QL LAI+AVF SW N RA YR+++ I E LG

Sbjct: 181 LKKLVERYKQIY-KEEGKEFPQDPWKQLWLDAVFGSWNNPRAKIYRQIHIGKEGDLG 239

25

Query: 235 TAVNIQAMVFGNMGNNSGTGVAFTRNPTGAANLFGYELINAQGEDVVAGIRTPQSISK 294

TAVNI AMVFGNMG +SGTGVAFT+P+TG +GE+L NAQGEDVVAGIRTP + +L

Sbjct: 240 TAVNIVAMVFGNMGDSGTGVAFTRDPTNGEKKPYGEFLPNAQGEDVVAGIRTPLEEL 299

30

Query: 295 AEQMPIIYQEFVSVTQKLEAHYRDMQDMEFTIENGNYMLQTRSGKRTAKAAIKIAVDQV 354

+MP +Y + + + KLE HYRDMQD+EFT+E G LY+LQTR+GKRT++AAI+IAVD V

Sbjct: 300 KNRMPDEVYNQLEIMDKLEKHYRDMQDIEFTVERGKLYILQTRNGKRTSQAIRIAVDMV 359

35

Query: 355 NEGLISKEEAILRIEPKQLDQLLHPSFDLKSLLKAIILTGLPASPGAAAYGVYFHAEDV 414

+EGLI+KEEAILR+ P+ ++Q+LHP FD K +A ++ GLPASPGAA GKV F+A+

Sbjct: 360 HEGITKEEAILRVRPEDVEQVILHPVDFPKEKAQAKVIAGLPGASPGAAATGKVVFNAKKA 419

40

Query: 415 VKEMKGNPVLLVRQETSPEDIEGMVSANGIITARGGMTSHAAVVARGMGKPCVAGCSQL 474

+ K G V+LVR ETSPE+ GM +A GI+T+RGGMTSHAAVVARGMGK V G +

Sbjct: 420 EELGKAGEQVILVRPETSPEVGGMAAAQGILTSRGGMTSHAAVVARGMGKPAVVGAESE 479

45

Query: 475 LVDEVRRREISIGHQTIKEGEMLSIDGATGNVYIGQV-PMAETSVD R DFEIFMKWVDENRD 533

V +G +KEGE +SIDG TG V +G+V + ++ ++W DE R

Sbjct: 480 EVHPEEGYFKVGDVVVKEGEWISIDGTTGEVLLGKVITIKPQGLEGPVAELLQWADEIRR 539

50

Query: 534 MMVCSNADNPRDAQKALDFGAEIGLCRTEHMFDDERIPVVREMILADEILSRKALER 593

+ V +NAD PRDA+ A FGAEGIGLCRTEHMF+ +RIP VR MILA R KAL+

Sbjct: 540 LGVRTNADIPRDAEVARKFGAEGIGLCRTEHMFEEKDRIPKVRMILAKTKEEREKALDE 599

55

Query: 594 LLSFQRDDFYQIFKVLKGACTIONRLDPPLHEFLPHDKESIESMARQMGISTLAIEKRIQ 653

LL Q++DF +F+V+KG TIRL+DPPLHEFLP + E I+ +A QMG+S ++ ++

Sbjct: 600 LLPLQKEDFKGLFRVMKGLPVITIRLIDPPLHEFLPQEDQIKEVAEQMGVSFEELKNVVE 659

60

Query: 654 TLEEFNPMLGHRGCRLLAITYPEIQMVQVRLVQGA-LAMKEGYEAKPEIMIPLVTAHHE 712

L+E NPMLGHRGCRLL IYPEI MQ +A++ AI L +EG + PEIMIPLV E

Sbjct: 660 NLKEINPMLGHRGCRLLIITYPEIAVMQTKAIIIGAAIELKKEEGIDVPEIMIPLVGHVNE 719

65

Query: 713 ISIIRDLIEETIVEESKSKKINLSFPIGTMIETPRACMIADDIKAFADFFSFGTNDLTQM 772

+ ++ +I+ET K + L++ IGTMIET PRA + A IA+ A+FFSFGTNDLTQM

Sbjct: 720 LRYLKKIIEKETADALIKEAGVELTYKIGTMIEVPRAAVTAHQIAEEAEFFSFGTNDLTQM 779

Query: 773 SFGFSRDDAGKFLGEYVDKGLLKKDPFQVLDQKGIGRFIGQAVRLGKEVKPNLKIGICGE 832

+FGFSRDD GKFL EY++KG+L+ DP+ LD G+G + G+ +P+LK+G+CGE

Sbjct: 780 TFGFSRDDVGKFLPEYLEKGILEDPFKTLDYDGVGELVRMGKEKGRSTRPDLKVGVCGE 839

Query: 833 HGGEPSSIEFCYQLGLHYVSCSPFRIPIAKLAAAQAKIKQSR 874

HGG+P SI F ++GL YVSCSP+R+P+A+LAAAQA +K +

Sbjct: 840 HGGDPSRILFFDKIGLDYVSCSPYRVPVARLAAAQAALKNKK 881

-2028-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1793

A DNA sequence (GBSx1900) was identified in *S.agalactiae* <SEQ ID 5573> which encodes the amino acid sequence <SEQ ID 5574>. This protein is predicted to be glutamyl-tRNA (Gln) amidotransferase subunit C (gatC). Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04384 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
subunit C [Bacillus halodurans]

Identities = 42/94 (44%), Positives = 63/94 (66%)

Query: 2 KISEEEVRHVNLSKLRFSQETKEFASSLSKIVDMIELLNEVDTEGVPVTTTMADRKT 61
+IS E+V+HVA+L++L +++E K F L I+ E LNE+DTEGV T+ + D K V
Sbjct: 3 RISMEQVKHVAHLARLAITEEAKLFTEQLGDIIQFAEQLNELDTEGVEPTSHVLDMKNV 62

Query: 62 MREDIAQPGHNRDDL FKNVPQHQQDYIYKVPAIL 95
+RED + G +D+ KN P H+D I+VP++LE
Sbjct: 63 LREDKPEKGLPVEDVLKNAPDHEDGQIRVPSVLE 96

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5575> which encodes the amino acid sequence <SEQ ID 5576>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3247(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 72/100 (72%), Positives = 88/100 (88%)

Query: 1 MKISEEEVRHVNLSKLRFSQETKEFASSLSKIVDMIELLNEVDTEGVPVTTTMADRKT 60
MKISEEEVRHVA LSKL FS+ ET FA++LSKIVDM+ELLNEVDTEGV +TTTMAD+K
Sbjct: 5 MKISEEEVRHVAKLSKLSFSESETTTTATTLISKIVDMVELLNEVDTEGVAITTTMADKKN 64

Query: 61 VMREDIAQPGHNRDDL FKNVPQHQQDYIYKVPAILDGGDA 100
VMR+D+A+ G +R LFKNVP+ ++++IKVPAIL+DGGDA
Sbjct: 65 VMRQDVAEEGTDRALLFKNVPEKENHFIVKVPAILDDGGDA 104

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1794

A DNA sequence (GBSx1901) was identified in *S.agalactiae* <SEQ ID 5577> which encodes the amino acid sequence <SEQ ID 5578>. Analysis of this protein sequence reveals the following:

-2029-

Possible site: 30

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -7.64 Transmembrane 7 - 23 (6 - 24)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.4057(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1795

15 A DNA sequence (GBSx1902) was identified in *S.agalactiae* <SEQ ID 5579> which encodes the amino acid sequence <SEQ ID 5580>. This protein is predicted to be glutamyl-tRNA amidotransferase, subunit A (gatA). Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

20

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2855(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04385 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase

subunit A [Bacillus halodurans]

Identities = 285/486 (58%), Positives = 367/486 (74%), Gaps = 4/486 (0%)

30

Query: 1 MSFNNQSIDQLHDFLVKKEISATELTATLEDIHAREQAVGSFITISDEMAIAQAKEID- 59

MS + + +H L +KEIS ++L + I + V +F+ +++E A A AKE+D

Sbjct: 1 MSLFDLKLKDVHTKLHEKEISVSDLVDEAYKRIEQVDGQVEAFALNEEKARAYAKELDA 60

35

Query: 60 --DKGIDADNVMSGIPLAVKDNISTKGILTTAASKMLYNEPIFDATAVEKLYAKDMIVI 117

D+ +A ++ GIP+ VKDNI TK + TT +S++L N++PI+DAT V KL + I

Sbjct: 61 ALDRS-EARGLLFGIPIGVKDNIVTKNLRITTCSSRILGNFDPIDYDATVVKHLREAQAVTI 119

40

Query: 118 GKANMDEFAMGGSTETSYFKKTNNAWDHSKVPGGSSGGSAASAVASGQVRLSLGSDTGGSI 177

GK NMDEFAMG STE S F+KT N W+ VPGSSGGSAASAVA+G+V +LGSDTGGSI

Sbjct: 120 GKLNDEFAMGSSSTENSAFQKTKPNWNLEYVPGSSGGSAASAVAAGEVPFTLGSDTGGSI 179

45

Query: 178 RQPASFNIGIVGMKPTYGRVSRFGLFAFGSSLDQIGPMSQTVKENAQLLTIVISGHDVRDST 237

RQPA++ G+VG+KPTYGRVSR+GL AF SSLDQIGP+++ V++NA LL ISGHD DST

Sbjct: 180 RQPAAYCGVVGLKPTYGRVSRVGLVAFASSLDQIGPITRNVEDNAYLLQAISGHDPMDDST 239

50

Query: 238 SSERTVGDFTAKIGQDIQGMKIALPKEYLGEGLAQGVKETIIKAAKHLEKLGAVIEEVSL 297

S+ V D+ + + DI+G+KIA+PKEYLGEGL+ + VK++++ A K LE LGA EEVSL

Sbjct: 240 SANLDVPDYLSTALTDIKGLKIAVPKEYLGEGLVKEEVKQSVLDALKVLEGLGATWEEVSL 299

55

Query: 298 PHSKYGVAVYYIVASSEASSNLQRFDGIRYGYRTENYKNLDDIYVNTTRSEFGDEVKRRRI 357

PHSKY +A YY++ASSEAS+NL RFDG+RYG+R++N NL D+Y TR+EGFGDEVKRRRI

Sbjct: 300 PHSKYALATYYLLASSEASANLARFDGVRYGFRSDNADNLLDMYKQTRAEGFGDEVKRRRI 359

Query: 358 MLGTFSLSSGYDAYYYKKAQGVRLIIQDFEKVFADYDLILGPTAPTAFDLDLSLNHDPV 417

MLGTF+LSSGYDAYYYKKA QVR+LI QDFEKVF YD+I+GPT PT AF + DP+

Sbjct: 360 MLGTFALSSGYDAYYYKKAQQVRLTIKQDFEKVFQYDVIIGPTTPTPAFKIGEKTDDEPL 419

Query: 418 AMYLADILTIPVNLAGLPGISIPAGFDQGLPVGMQLIGPKFSEETIYQVAAFEATTDYH 477

-2030-

MY DILTIPVNLG+P IS+P GFD GLP+G+Q+IG F E ++Y+VA AFE TDYH
 Sbjct: 420 TMYANDILTIPVNLGVP AISVPCGFDNGLPLGLQIIGKHFDGSGVYRVVAHAFEQATDYH 479

Query: 478 KQQPKI 483

++P +

Sbjct: 480 TKRPTL 485

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5581> which encodes the amino acid sequence <SEQ ID 5582>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 392/487 (80%), Positives = 442/487 (90%)

Query: 1 MSFNQSIDQLHDFLVKKEISATELTATLEDIHAREQAVGSFITISDEMAIAQAKEIDD 60

MSFN+++I++LHD LV KEISATELT+ATLEDI +RE+AVGSFIT+S+E+A+ QA ID

Sbjct: 1 MSFNHKTIEELHDLVLAKEISATELTQATLEDIKSREEAVGSFITVSEEVALKQAAIDA 60

Query: 61 KGIDADNVMSGIPLAVKDNISTKGILTTAASKMLYNYEPIFDATAVEKLYAKDMIVIGKA 120

KGIDADN+MSGIPLAVKDNISTK ILTTAASKMLYNYEPIF+AT+V YAKDMIVIGK

Sbjct: 61 KGIDADNLMSGIPLAVKDNISTKEILTTAASKMLYNYEPIFNATSVANAYAKDMIVIGKT 120

Query: 121 NMDEFAMGGSTETSYFKKTNNAWDHKVPGGSSGGSAAAVASGQVRLSLGSDTGGSIQRP 180

NMDEFAMGGSTETSYFKKT NAWDH+KVPGGSSGGS AAVASGQVRLSLGSDTGGSIQRP

Sbjct: 121 NMDEFAMGGSTETSYFKKTNNAWDHKTVPGGSSGGSATAVASGQVRLSLGSDTGGSIQRP 180

Query: 181 ASFNGIVGMKPTYGRVSRFGLFAGSSSLDQIGPMSQTVKENAQLLTVISGHDVRDSTSSE 240

A+FN +VG+KPTYG VSR+GL AFGSSSLDQIGP + TVKENAQLL VI+ DV+D+TS+

Sbjct: 181 AAFNSVVGLKPTYGTVSRYGLIAGSSSLDQIGPFAPTQVKENAQLLNVIASSDVKDATSAP 240

Query: 241 RTVGDFATAKIGQDIQGMKIALPKEYLGEGIAQGVKETI IKAHLEKLGAVIEEVSLPHS 300

+ D+T+KIG+DI+GMKIALPKEYLGEGI +KET++ + K E LGA +EEVSLPHS

Sbjct: 241 VRIADYTSKIGRDIKGMKIALPKEYLGEGIDPEIKETVLA SVKQFEALGATVEEVSLPHS 300

Query: 301 KYGVAVYYYIVASSEASSNLQRFDGIRYGYRTENYKNLDDIYVNTRSEGFGDEVKRRIMLG 360

KYGVAVYYYI+ASSEASSNLQRFDGIRY+R ++ KNLD+IYVNTRS+GFGDEVKRRIMLG

Sbjct: 301 KYGVAVYYYIIASSEASSNLQRFDGIRYGFRAADAKNLDEIYVNTRSQFGDEVKRRIMLG 360

Query: 361 TFSLSGGYYDAYYKKAGQVRSLLIQDFEKFVADYDLILGPTAPTAFDLDSLNDHPVAMY 420

TFSLSGGYYDAY+KKAGQVR+LLIQDF+KVFADYDLILGPT PT AF LD+LNHDPVAMY

Sbjct: 361 TFSLSGGYYDAYYKKAGQVRTLLIQDFDKVFADYDLILGPTTPTVAFGLDTLNHDPVAMY 420

Query: 421 LADILTIPVNLGAGLPGISIPAGFDQGLPVGMQLIGPKFSEETIYQVAAAFEATTDYHKQQ 480

LAD+LTIPVNLGAGLPGISIPAGF GLPVG+QLIGPK++EETIYQ AAAFEA TDYHKQQ

Sbjct: 421 LADLLTIPVNLGAGLPGISIPAGFVGLGLQIGPKYAEETIYQAAAAFEAVTDYHKQQ 480

Query: 481 PKIFGGE 487

P IFGG+

Sbjct: 481 PIIFGGD 487

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2031-

Example 1796

A DNA sequence (GBSx1903) was identified in *S.agalactiae* <SEQ ID 5583> which encodes the amino acid sequence <SEQ ID 5584>. This protein is predicted to be glutamyl-tRNA^{Gln} amidotransferase subunit B (gatB). Analysis of this protein sequence reveals the following:

```

5   Possible site: 27
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10095> which encodes amino acid sequence <SEQ ID 10096> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB04386 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
      subunit B [Bacillus halodurans]
      Identities = 308/476 (64%), Positives = 361/476 (75%), Gaps = 1/476 (0%)

20 Query: 1  MNFETVIGLEVHVELNTNSKIFSPSSAHFGQEQNANTINVIDWSFPGVLPVMNKGVIDAGI 60
      MNFETVIGLEVHVEL T SKIFS S HFG E NANT+VID +PGVLPV+NK ++ +
      Sbjct: 1  MNFETVIGLEVHVELKTESKIFSASPNHFGAEPNANTSVIDLGYPGVLPVLNKA AVEFAM 60

Query: 61  KAALALNMDIHQNMHFDKKNFYFDPNPKAYQISQFDEPIGYNGWIEIELEDGTRKKIRIE 120
25 KAA+ALN ++ + FDRKNFYFDPNPKAYQISQFD+PIG NGWIEIE+ DGT+KKI I
      Sbjct: 61  KAAMALNCEVATDTKFDKKNFYFDPNPKAYQISQFDKPIGENGWIEIEV-DGTRKKKIGIT 119

Query: 121  RAHLEEDAGKNTHTGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180
30 R HLEEDAGK TH +GYS VD NRQG PLIEIVSE D+R+P+EAYAYL LK IIQYTG
      Sbjct: 120  RLHLEEDAGKLTHTSGNGYSLVDFNRQGTPLIEIVSEPDIRTPQEAYAYLEKLKSIIQYTG 179

Query: 181  ISDVKMEEGSMRVDANISLRPYGQEEFGTKAELKNLNSFNVRKGLIHEEKRAQVLRSG 240
      +SD KMEEGS+R DANISLRP QQEEFGTK ELKNLNSFN VRKGL +EEKRAQVLR SG
35 Sbjct: 180  VSDCKMEEGSLRCDANISLRPVGQEEFGTKTELKNLNSFNVRKGLYEYEEKRAQVLLSG 239

Query: 241  GOIQQETRRFDETTGETILMRVKEGSSDYRYFPEPDLPLFDISDEWIDQVRLELPEFPQE 300
      G+I QETRR+DE +T+LMRVKEGS DYRYFPEPDL I DEW ++R E+PE P
40 Sbjct: 240  GEILQETRRYDEAANKTVLMRVKEGSDDYRYFPEPDLVALHIDDEWKARIRSEIPELPDA 299

Query: 301  RRAKYVSSFGLSSYDASQLTATKATSDFFEKAVAIGGDAKQVSNWLQGEVAQFLNSESKE 360
      R+ +YV GL +YDA LT TK SDFPE+ +A G D K SNWL GEV+ +LN+E K
45 Sbjct: 300  RKKRYVEELGLPAYDAMVLTLTKEMSDFFEETIAKGADPKLASNWLNGEVSGYLNAEQKE 359

Query: 361  IEEIGLTPENLVEMIGLIADGTISSKIAKKVFVHLAKNGGSAEEFVKKAGLVQISDPEVL 420
      ++E+ LTP+ L +MI LI GTISSKIAKKVF L + GG EE VK GLVQISD L
50 Sbjct: 360  LDEVALTPDGLAKMIQLIEKGTISSKIAKKVFKDLIEKGGDPPEIVKAGLVQISDEGEL 419

Query: 421  IPIIHQVFADNEAAVIDFKSGKRNADKFTGYLMKATKGQANPQVALKLLAQELAK 476
      + +V +N+ ++ DFK+GK A G +MKATKG+ANP + KLL +E+ K
      Sbjct: 420  RKYVVEVLDDNNQQSIDDFTKNGKDRAGFLVGQIMKATKGKANPPMVNKLILLEINK 475

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5585> which encodes the amino acid sequence <SEQ ID 5586>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 27
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
60      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2032-

An alignment of the GAS and GBS proteins is shown below.

Identities = 410/479 (85%), Positives = 447/479 (92%)

```

5  Query: 1  MNFETVIGLEVHVELNTNSKIFSPSSAHFGQEQNANTNVIDWSFPGVLPVMNKGVIDAGI 60
   Sbjct: 1  MNFET+IGLEVHVELNTNSKIFSPSSAHFG++ NANTNVIDWSFPGVLPVMNKGVIDAGI 60

10 Query: 61  KAALALNMDIHQNMFDRKNFYFDPNPKAYQISQFDEPIGYNGWIEIELEDGTRKKIRIE 120
   Sbjct: 61  KAALALNMDIH+ MHFDRKNFYFDPNPKAYQISQFDEPIGYNGWI+I+LEDG+ KKIRIE 120

15 Query: 121  RAHLEEDAGKNTHTGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180
   Sbjct: 121  RAHLEEDAGKNTHTGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180

20 Query: 181  ISDVKMEEGSMRVDANISLRPYGQEEFGTKAELKNLNSFNVRKGLIHEEKROAQVLRSG 240
   Sbjct: 181  ISDVKMEEGSMRVDANISLRPYGQE+FGTK ELKNLNSF+NVRKGL E +RQA++LRSG 240

25 Query: 241  GQIQQETRRFDETTGETILMRVKEGSSDYRYFPEPDLPLFDISDEWIDQVRLELPEFPQE 300
   Sbjct: 241  G I+QETRR+DE TILMRVKEG++DYRYFPEPDLPL++I D WID++R +LP+FP + 300

30 Query: 301  RRAKYVSSFGLSSYDASQLTATKATSDFFEKAVAIGGDAKQVSNWLQGEVAQFLNSESKS 360
   Sbjct: 301  RRAKY GLS+YDASQLTATK SDFFE AV++GGDAKQVSNWLQGEVAQFLN+E K+ 360

35 Query: 361  IEEIGLTPENLVEMIGLIADGTISSKIAKKVFVHLAKNGGSAEEFVKKAGLVQISDPEVL 420
   Sbjct: 361  IEEI LTPENLVEMI +IADGTISSK+AKKV FVHLAKNGGSA +V+KAGLVQISDP VL 420

   Query: 421  IPIIHQVFADNEAAVIDFKSGKRNADKAFTGYLMKATKGQANPQVALKLLAQELAKLKE 479
   Sbjct: 421  VPIIHQVFADNEAAVADFKSGKRNADKAFTGFLMKATKGQANPQVAQLLAQELQKLRLD 479

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1797

40 A DNA sequence (GBSx1904) was identified in *S.agalactiae* <SEQ ID 5587> which encodes the amino acid sequence <SEQ ID 5588>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have an uncleavable N-term signal seq

```

45  INTEGRAL    Likelihood = -7.27    Transmembrane 108 - 124 ( 105 - 125)
   INTEGRAL    Likelihood = -7.27    Transmembrane 278 - 294 ( 268 - 301)
   INTEGRAL    Likelihood = -6.05    Transmembrane 191 - 207 ( 188 - 208)
   INTEGRAL    Likelihood = -5.63    Transmembrane 219 - 235 ( 215 - 242)
   INTEGRAL    Likelihood = -3.93    Transmembrane 41 - 57 ( 39 - 58)
50  INTEGRAL    Likelihood = -3.88    Transmembrane 132 - 148 ( 131 - 150)
   INTEGRAL    Likelihood = -3.03    Transmembrane 254 - 270 ( 253 - 272)
   INTEGRAL    Likelihood = -3.03    Transmembrane 79 - 95 ( 79 - 95)

```

----- Final Results -----

```

55  bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10093> which encodes amino acid sequence <SEQ ID 10094> was also identified.

60 The protein has homology with the following sequences in the GENPEPT database.

-2033-

>GP:CAA04271 GB:AJ000733 hypothetical protein [Bacillus megaterium]
Identities = 102/292 (34%), Positives = 169/292 (56%), Gaps = 3/292 (1%)

5 Query: 6 TKKEKGTMMTLAAGLAWGISGSGQYLMHS-GVHVNLITSLRLLITGIFLLSLARSKQKE 64
+++ G ++ + WG+SG QYL H + L +R+L++G+ LL++A SKQ+
Sbjct: 1 SRRAWGLLLVIIGATMWGVSQTVAQYLFQHKSEFAEWLVVVRMLVSGLLLLAIA-SKQR- 58

10 Query: 65 HLVAAWKQPKFLKQVLLFSIFGLVLNQYAFRLAIHLTNAGTATVLQYMAPILILSIVCIL 124
++ A WK + +LLF + G++ QY + AI NA TATVLQY +PI I+ + +
Sbjct: 59 NIFAIWKTEERTSLLLFGVIGMLGVQYTYFAAIEAGNAATATVLQYTSPIFIIGYLAVQ 118

15 Query: 125 NRQRPTSFEIIATAMAILGTMYMIATHGKLGSLAITPKGLMWGLGSAITYSIVILLPVKLI 184
R+ P E+I++ + I GT+ +AT G L+IT L WG+G+A+T + Y L P +L+
Sbjct: 119 ARKWPVKVEMISVVLVIAGTFFLATSGNFNELSGITGWALFWGIGAAVTSAFYTLQPKRLL 178

20 Query: 185 HEWGSTIVIGSGMFIGGILFSLVTKAWQYPLQINVMSILAYIGIIGTIFAYTFFLKGV 244
+W S V+G GM IGG FS + W + +++S+ A + +I GT+ A+ +L+ +
Sbjct: 179 AKWSSIEVVWGMVIGGASFSFIHPPWHIAGEWSLLSLCAVLFVIIFGTLIAFYCYLES 238

25 Query: 245 SIVGAVKGSLLASVEPVSSVFLTVLVLGEIFYPIDLLGMLFIFLAVTLISYK 296
+ A + +LAS EP+S+ L+VL L F + LG + I V L+S +
Sbjct: 239 KHISASEAIVLASREPLSAAALSVLWLHVTFGWTEWLGTLIIATVFLLSQR 290

No corresponding DNA sequence was identified in *S.pyogenes*.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1798

A DNA sequence (GBSx1905) was identified in *S.agalactiae* <SEQ ID 5589> which encodes the amino acid sequence <SEQ ID 5590>. Analysis of this protein sequence reveals the following:

30 Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35 bacterial cytoplasm --- Certainty=0.2103(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10091> which encodes amino acid sequence <SEQ ID 10092> was also identified.

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14510 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
Identities = 52/153 (33%), Positives = 88/153 (56%)

45 Query: 17 YRPTFVVEAVYDLRAEDLLRHGIRAVLVDLDNTLIAWNNPDGTAEVRAWLDEMTTADISV 76
+ P V+ ++ + E L ++ ++ DLDNTL+ W+ P+ T + W +EM I V
Sbjct: 6 FLPDFVKNIFHITPEKLKERNVKGIIITDLDNTLVEWDRPNATPRLEWFEEEMKEHGIKV 65

50 Query: 77 VVVSNNNHARVERAVSRFGVDFVSRAMKPFTRGINMAIERYGFDRDEVIMVGDQLMTDIR 136
+VSNNN RV+ G+ F+ +A KP + N A+ +++ ++GDQL+TD+
Sbjct: 66 TIVSNNNERRVKLFSEPLGIPFTYKARKPMGKAFNRRAVRNMELKKEDCVVIGDQLLTDVL 125

55 Query: 137 ASHRAGIKSVLVKPIVKSDAUNTGFNRLRERRV 169
+R G ++LV P+ SD + T+FNR ERR+
Sbjct: 126 GGNRNGYHTILVVPVASSDGFITRFNRQVERRI 158

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5591> which encodes the amino acid sequence <SEQ ID 5592>. Analysis of this protein sequence reveals the following:

-2034-

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4252(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 147/175 (84%), Positives = 158/175 (90%)

Query: 12 LSIDDYRPTFFVVEAVYDLRAEDLLRHGIRAVLVLDNTLIAWNNPDGTAEVRAWLDEMTT 71
 +SIDDDYRPT++VEA+YDLRA DLLRHGI AVLVDLDNTLIAWNNPDGT EVRAWLDEMT
 Sbjct: 20 MSIDDYRPTYMVEAIYDLRANDLLRHGITAVLVLDNTLIAWNNPDGTPEVRAWLDEMTI 79

15 Query: 72 ADISVVVVSNNNHARVERAVSRFGVDFVSRAMKPFTRGINMAIERYGFDRDEVIMVGDQL 131
 ADISVVVVSNN H+RVERAVSRFGVDF+SRA+KPF GI AI RYGFDR+EVIMVGDQL
 Sbjct: 80 ADISVVVVSNNNKHSRVERAVSRFGVDFISRALKPFAYGIEKAIARYGFDRNEVIMVGDQL 139

20 Query: 132 MTDIRASHRAGIKSVLVKPIVKSDAWNTKFNRLRERRVWKKIEENYGKIVYQKGI 186
 MTDIRASHRAGIKSVLVKP+V SDAWNTK NR RERRV K+EE YGK+ YQKGI
 Sbjct: 140 MTDIRASHRAGIKSVLVKPLVASDAWNTKINWRERRVMAKLEEKYKLSYQKGI 194

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1799

A DNA sequence (GBSx1906) was identified in *S.agalactiae* <SEQ ID 5593> which encodes the amino acid sequence <SEQ ID 5594>. Analysis of this protein sequence reveals the following:

30 Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.1091(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14509 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 220/373 (58%), Positives = 280/373 (74%), Gaps = 8/373 (2%)

40 Query: 1 MEELFCIGCGARIQTENKDAAGYTPRAALEKGLETGELYCQRCFRLRHYNIEITDVHITDD 60
 ME++ CIGCG IQTE+K GY P A+L K + CQRCFRL++YNEI DV +TDD
 Sbjct: 1 MEKVVICIGCVTIQTEDKTGLGYAPPASLTKE---NVICQRCFRLKNYNEIQDVSLTDD 56

45 Query: 61 EFLKLLHEVGDS DALVVNVIDIFDFNGSIIPGLSRFVAGNDVLLVGNKKDILPKSVKDGK 120
 +FL +LH +G++D+LVV ++DIFDFNGS I GL R V GN +LLVGNK DILPKS+K +
 Sbjct: 57 DFLNILHGIGETDSL VVKIVDIFDFNGSWINGLQRLVGNPILLVGNKADILPKSLKRER 116

50 Query: 121 VTQWLTERAHEEGLRPVDVILTSAQNHHAIKDLIDTIEKYRHGQDVYVVGVTNVGKSTLI 180
 + QW+ A E GL+PVDV L SA I+++ID IE YR+G+DVYVVG TNVGKST I
 Sbjct: 117 LIQWMKREAKELGLKPVDVFLVSAGRGQIREVIDAIEHYRNGKDVYVVGCTNVGKSTFI 176

55 Query: 181 NAIIREITGSRDVITTSRFPGTTLDKIEIPLDDGSYIFDTPGIIHRHQMAHYLTAKNLKY 240
 N II+E++G D+ITTS+FPGTTL D IEIPLDDGS ++DTPGII+ HQMAHY+ K+LK
 Sbjct: 177 NRIIKEVSGEEDIITTSQFPGTTLDAIEIPLDDGSSLYDTPGIIINNMQMAHYVNKKDLKI 236

60 Query: 241 VSPKKEIKPKTYQLNSEQTLFLAGLARFD FISGKQGF TAYFDNNLNLHRTKLVGADEFY 300
 +SPKKE+KP+T+QLN +QTL+ GLARFD++SG++ F Y N L +HRTKL AD Y
 Sbjct: 237 LSPKKELKPRTFQLNDQQTLYFGGLARFDYVSGERSPFICYMPNLMIHRTKLENADALY 296

Query: 301 TKHV GKLLTPPTGKEVSDFPKLV RHEFTIKD-KMDIVYSGLGWIRVKSEAENPVVVAWA 359

-2035-

KH G+LLTPP E+ +FP+LV H FTIKD K DIV+SGLGW+ V + V A+A
 Sbjct: 297 EKHAGELLTPPGKDEMEFFPELVHFTTIKDKKTDIVFSGLGWVTVHDADKK---VTAYA 353

Query: 360 PEGVAVVLRKALI 372

P+GV V +R++LI

Sbjct: 354 PKGVHVFVRRSLI 366

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5595> which encodes the amino acid sequence <SEQ ID 5596>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14509 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 220/373 (58%), Positives = 286/373 (75%), Gaps = 8/373 (2%)

Query: 1 MEELFCIGCGIQTEDEKAGFTPAALKKGMETGELYCQRCFRLRHYNEITDVHITDD 60

ME++ CIGCG+ IQTEDEK G+ P A+L K + CQRCFRL++YNEI DV +TDD

Sbjct: 1 MEKVVCIGCGVTIQTEDEKGLGYAPPASLTKE----NVICQRCFRLKKNYNEIQDVSLTDD 56

Query: 61 EFLRLLEHVGDS DALVVNVIDIFDFNGSIIPGLSRFISGNDVLLVGNKKDILPKSVKDGK 120

+FL +LH +G++D+LVV ++DIFDFNGS I GL R + GN +LLVGNK DILPKS+K +

Sbjct: 57 DFLNHLHGIGETDSL VVKIVDIFDFNGSWINGLQRLVGNPILVGNKADILPKSLKRER 116

Query: 121 VTQWLTERAHEEGLRPLDVMLTSAQNKYAIKDLIGRINELRNGRDVYVVGVTNVGKSTLI 180

+QW+ A E GL+P+DV L SA I+++I I RNG+DVYVVG TNVGKST I

Sbjct: 117 LIQWMKREAKELGLKPVDFLVSAAGRGQIREVIDAIEHYRNGKDVYVVGCTNVGKSTFI 176

Query: 181 NAIIEITGNKDVITTSRFPGTTLDKIEIPLDDGTIFDTPGIIHRHQMAHYLSPKELKI 240

N II+E++G +D+ITTS+FPGTTL D IEIPLDDG+ ++DTPGII+ HQMAHY++ K+LKI

Sbjct: 177 NRIIEKVSGEEDIITTSQFPGTTLDAIEIPLDDGSSLYDTPGIINN HQMAHYVNKKDLKI 236

Query: 241 VSPKKEIKPKTYQLNPEQTLFLGLARFDFINGERQGFATFDNQLELHRTKLAGADAFY 300

+SPKKE+KP+T+QLN +QTL+ GGLARF+++GER F + N+L +HRTKL ADA Y

Sbjct: 237 LSPKKELKPRTFQLNDQQTLYFGGLARFDYVSGERSPFICYMPNEMIHR TKLENADALY 296

Query: 301 DKHVGTLTPPDKKELTAFPKLVRHEFTI-DQKMDIVFSGLGWIRVNGQKDSKAIVA AWA 359

+KH G LLTPP K E+ FP+LV H FTI D+K DIVFSGLGW+ V+ D+ V A+A

Sbjct: 297 EKHAGELLTPPGKDEMEFFPELVHFTTIKDKKTDIVFSGLGWVTVH---DADKKVTAYA 353

Query: 360 PEGVAVIVRKAI 372

P+GV V VR+++I

Sbjct: 354 PKGVHVFVRRSLI 366

An alignment of the GAS and GBS proteins is shown below.

Identities = 308/372 (82%), Positives = 343/372 (91%)

Query: 1 MEELFCIGCGARIQTENKDAAGYTPRAALEKGLTGLYQRCFRLRHYNEITDVHITDD 60

MEELFCIGCG +IQTE+K+ AG+TP AAL+KG+ETGELYCQRCFRLRHYNEITDVHITDD

Sbjct: 1 MEELFCIGCGIQTEDEKAGFTPAALKKGMETGELYCQRCFRLRHYNEITDVHITDD 60

Query: 61 EFLKLLHEVGDS DALVVNVIDIFDFNGSIIPGLSRFVAGNDVLLVGNKKDILPKSVKDGK 120

EFL+LLHEVGDS DALVVNVIDIFDFNGSIIPGLSRF++GNDVLLVGNKKDILPKSVKDGK

Sbjct: 61 EFLRLLEHVGDS DALVVNVIDIFDFNGSIIPGLSRFISGNDVLLVGNKKDILPKSVKDGK 120

Query: 121 VTQWLTERAHEEGLRPVDVILTSAQNHHAIKDLIDTIEKYRHGQDVYVVGVTNVGKSTLI 180

VTQWLTERAHEEGLRP+DV+LTS AQN +AIKDLI I + R+G+DVYVVGVTNVGKSTLI

Sbjct: 121 VTQWLTERAHEEGLRPLDVMLTSAQNKYAIKDLIGRINELRNGRDVYVVGVTNVGKSTLI 180

-2036-

Query: 181 NAIIREITGSRDVITTSRFP GTTLDKIEIPLDDGSYIFDTPGIIHRHQAHYLTAKNLKY 240
 NAIIEITG++DVITTSRFP GTTLDKIEIPLDDG++IFDTPGIIHRHQAHYL+ K LK
 Sbjct: 181 NAIIEITGNKDVITTSRFP GTTLDKIEIPLDDGTFI DTPGIIHRHQAHYLSPKELKI 240

Query: 241 VSPKKEIKPKTYQLNSEQTFLFLAGLARFD FISGQKGFTAYFDNNLNLHRTKLVGADEFY 300
 VSPKKEIKPKTYQLN EQTLFL GLARFDI+G++QGFTA+FDN L LHRTKL GAD FY
 Sbjct: 241 VSPKKEIKPKTYQLNPEQTFLFLGLARFD FINGERQGF TAFFDNQLELHRTKLAGADAFY 300

Query: 301 TKHVGKLLTPPTGKEVSDFPKLV RHEFTIKDKMDIVYSGLGWIRVKSEAENPVVVAWAP 360
 KHVG LLTPP KE++ FPKLV RHEFTI KMDIV+SGLGWIRV + ++ +VAWAP
 Sbjct: 301 DKHVG TLLTPPDKKELTAFPKLV RHEFTIDQKMDIVYSGLGWIRVNGQKDSKAIVA WAP 360

Query: 361 EGVAVVLRKALI 372
 EGVAV++RKA+I
 Sbjct: 361 EGVAVIVRKAL I 372

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1800

A DNA sequence (GBSx1907) was identified in *Sagalactiae* <SEQ ID 5597> which encodes the amino acid sequence <SEQ ID 5598>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2948(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14507 GB:Z99117 similar to dihydrodipicolinate reductase

[Bacillus subtilis]

Identities = 49/97 (50%), Positives = 67/97 (68%), Gaps = 2/97 (2%)

Query: 1 MLTSKQRAFLKSEAHSMPKPIIQIGKNGLNDQIKTSVRNALDARELIKVTLLQNTDEDIHD 60

MLT KQ+ FL+S+AH + PI Q+GK G+ND + + AL+ARELIKV++LQN +ED +D

Sbjct: 1 MLTGKQKRFLRSKAHHLTPFQVGKGVNDNMKQIAEALARELIKVSVLQNCCEEDKND 60

Query: 61 VAEVLEDEIGCDTVLKIGRILILYKESARKENRKISV 97

VAE L V IG ++LYKES KEN++I +

Sbjct: 61 VAEALVKGSRSQLVQTIGNTIVLYKES--KENKQIEL 95

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5599> which encodes the amino acid sequence <SEQ ID 5600>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2839(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 89/102 (87%), Positives = 98/102 (95%)

Query: 1 MLTSKQRAFLKSEAHSMPKPIIQIGKNGLNDQIKTSVRNALDARELIKVTLLQNTDEDIHD 60

MLTSKQRAFLKSEAHS+KPI+QIGKNGLND IKTS+R ALDARELIKVTLLQNTDEDIH+

Sbjct: 1 MLTSKQRAFLKSEAHS LKPIVQIGKNGLNDHIKTSIRQALDARELIKVTLLQNTDEDIHE 60

-2037-

Query: 61 VAEVLEDEIGCDTVLKIGRILILYKESARKENRKISVKVKAV 102
 VAE+LE+EIGCDTVLKIGRILILYK SA+KENRK+S KVKA+
 Sbjct: 61 VAEILEEEIGCDTVLKIGRILILYKVSARKENRKLSPKVKAI 102

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1801

A DNA sequence (GBSx1908) was identified in *S.agalactiae* <SEQ ID 5601> which encodes the amino acid sequence <SEQ ID 5602>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 3 - 19 (1 - 21)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10089> which encodes amino acid sequence <SEQ ID 10090> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14506 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 85/187 (45%), Positives = 134/187 (71%)

Query: 38 KQIGIMGGNFNPVHNAHLVVADQVRQQLCLDQVLLMPEFQPPHIDKKETIDEQHRKMLE 97
 K+IGI GG F+P HN HL++A++V Q LD++ MP PPH ++ D HR++ML+
 Sbjct: 2 KKIGIFGGTFDPPHNGHLLMANEVLQAGLDEIWFMPNQIPPHKQNEDYTSFHRVEMLK 61

Query: 98 LAIEGIDGLSIEPIEIERKGISYTYDTMKLLIEKNPDVDYFIIGADMVEYLPKWHRIDE 157
 LAI+ +E +E+ER+G SYT+DT+ LL ++ P+ +FIIGADM+EYLPKW+++DE
 Sbjct: 62 LAIQSNPSFKLELVEMEREGPSYTFDTVSLKQRYPNQDLFFIIGADMIEYLPKWYKLDE 121

Query: 158 LVKMQVQFVGVRPKYKAGTSYPVIWDLPLMDISSMIRQFIKSNRQPNYLLPREVLDI 217
 L+ ++QF+GV+RP + T YP+++ D+P ++SS+MIR+ KS + +YL+P +V Y+
 Sbjct: 122 LLNLIQFIGVKRPGFHVETPYPLLFADVPEFEVSSTMIRERFKSKKPTDYLPDKVKKYV 181

Query: 218 RKEGLYK 224
 + GLY+

Sbjct: 182 EENGLYE 188

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5603> which encodes the amino acid sequence <SEQ ID 5604>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4660(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/210 (81%), Positives = 196/210 (92%)

Query: 15 MALELLTPFTKVELEEKKRDTNRKQIGIMGGNFNPVHNAHLVVADQVRQQLCLDQVLLMP 74
 MALELLTPFTKVELEE+K+++NRKQIGI+GGNFNP+HNAHLVVADQVRQQL LDQVLLMP
 Sbjct: 1 MALELLTPFTKVELEEKKESNRKQIGILGGNFNPIHNAHLVVADQVRQQLGLDQVLLMP 60

-2038-

Query: 75 EFQPPHIDKKETIDEQHRLKMLELAIEGIDGLSIEPIEIERKGISYTYDTMKLLIEKNPD 134
 E +PPH+D KETIDE+HRL+MLELAIE ++GL+IE E+ER+GISYTYDTM L E++PD
 Sbjct: 61 ECKPPHVDKETIDEKHRLRMLELAIEDVEGLAIEETCELERQGISYTYDTMLYLTEQHPD 120

Query: 135 VDYYFIIGADMVEYLPKWHRIDELVKMVQFVGVQRPKYKAGTSYPVIWVDLPLMDISSM 194
 VD+YFIIGADMV+YLPKWHRIDELVK+VQFVGVQRPKYKAGTSYPVIWVDLPL+DISSM
 Sbjct: 121 VDFYFIIGADMVDYLPKWHRIDELVKLVQFVGVQRPKYKAGTSYPVIWVDLPLIDISSM 180

Query: 195 IRQFIKSNRQPNYLLPREVLDYIRKEGLYK 224
 IR FIK RQPNYLLP+ VLDYI +EGLY+
 Sbjct: 181 IRDFIKKGRQPNYLLPKRVLDYITQEGLYQ 210

SEQ ID 5602 (GBS651) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 8-10; MW 53.3kDa) and in Figure 186 (lane 8; MW 53kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 12; MW 28.4kDa) and in Figure 140 (lane 11; MW 20kDa).

Purified GBS651-GST is shown in Figure 243, lane 4; purified GBS651-His is shown in Fig.229, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1802

A DNA sequence (GBSx1909) was identified in *S.agalactiae* <SEQ ID 5605> which encodes the amino acid sequence <SEQ ID 5606>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4281(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14505 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 79/180 (43%), Positives = 115/180 (63%)

Query: 9 LDRTELLSKVRHMSDKRFNHVLGVERAAIELAERYGYDKEKAGLAALLHDYAKELSDDE 68
 ++R E L+ V+ +++ R+ H +GV AIELAER+G D +KA +AA+ HDYAK +E
 Sbjct: 1 MNREEALACVKQQLTEHRYIHTVGMNTAIELAERFGADSKAEIAAIFHDYAKFRPKEE 60

Query: 69 FLRLIDKYQPDPLKKWGNNIWHGLVGIYKIQEDLAIKDQDILAAIAKHTVGSQMSTLD 128
 ++I + + L +WH VG Y +Q + ++D+DIL AI HT G M+ L+
 Sbjct: 61 MKQIIAREKMPAHLLDHNPPELWHAPVGAYLVQREAGVQDEDILDAIRYHTSGRPGMTLLE 120

Query: 129 KIVYVADYIEHNRDFPGVEEARELAKVDLNKAVAYETARTVAFLASKAQPTYPKTIETYN 188
 K++YVADYIE NR FPGV+E R+LA+ DLN+A+ T+ FL K QP++P T TYN
 Sbjct: 121 KVIYVADYIEPNRAFPVGVDEVKRLAETDLNQAALIQSIKNTMVFLMKKNQPVFPDFTLTYN 180

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5607> which encodes the amino acid sequence <SEQ ID 5608>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2615(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2039-

An alignment of the GAS and GBS proteins is shown below.

Identities = 130/194 (67%), Positives = 159/194 (81%)

```

5  Query: 1  MTKDYTGDLRTELLSKVRHMMSDKRFNHVLGVERAAIELAERYGYDKEKAGLAALLHDY 60
    MTY+DY  RTELL+K+  MS KRF HVLGVE+AA+ LAE YG + +KAGLAALLHDY
    Sbjct: 1  MTYEDYLPYSRTELLAKIAEQMSPKRFKHVLGVEKAALSLAECYGCNPDKAGLAALLHDY 60

10 Query: 61  AKELSDDEFLRLIDKYQDPDLKKWGNNIWHGLVGIYKIQEDLAIKDQDILAAIAKHTVG 120
    AKE D FL LIDKYQ P+L KW NN+WHG+VGIYKIQEDL +KD+DIL AI HTVG
    Sbjct: 61  AKECPDQVFLDLIDKYQLSPELAKWNNNVWHGMVGIYKIQEDLGLKDKDILRAIEIHTVG 120

    Query: 121  SAQMSTLDKIVYVADYIEHNRDFPGVEEARELAKVDLNKAVAYETARTVAFLASKAQPIY 180
    +A+M+ LDK++YVADYIE R FP V++AR++AK+DLN+AVAYET TVA+LASKAQPI+
15  Sbjct: 121  AAEMTLDDKVLVADYIEGRIFPLVDDARKIAKLDLNQAVAYETVNTVAYLASKAQPIF 180

    Query: 181  PKTIETYNYAYIPYL 194
    P+T++TYNA+ YL
20  Sbjct: 181  PQTLDTYNAFCSYL 194

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1803

A DNA sequence (GBSx1910) was identified in *S.agalactiae* <SEQ ID 5609> which encodes the amino acid sequence <SEQ ID 5610>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -2.34    Transmembrane    12 - 28.( 10 - 28)

30  ----- Final Results -----
        bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10087> which encodes amino acid sequence <SEQ ID 10088> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAG19496 GB:AE005041 Vng1100c [Halobacterium sp. NRC-1]
    Identities = 46/175 (26%), Positives = 82/175 (46%), Gaps = 12/175 (6%)

40  Query: 22  ALLLIDIQQGIMDKK--PKHLTNFAVLDDLLLSAKGSNCEVIWIRHHDKE----LPQGS 75
    AL+L+D QQG D ++ + ++LL + + + + +RH+ E L QG
    Sbjct: 7  ALVLVDFQQGFADPAWGDRNNPDAAEAHAELLAAWRDAAPIAHVRHNSTEATSPLRQGE 66

45  Query: 76  PQWEIWEQRHLVTHHKIIDKTYNSCFKDTLHLDYLSKHHISQLIMMGLQTEYCFDTSVKV 135
    P + + K+ N F DT L +L+ + L++ GL T++C T+V++
    Sbjct: 67  PGFAYTDGLAPAADEPEFVKSNGAFVDTALEGLWRDRDTGSLVVCGLTGDHCVSTTVRM 126

    Query: 136  AFEYGYDIFIPQGGHLTFDTPILSGDSIKK---HYENIWHHR--FATMVAKDSLL 185
    A G+D+ + + T D TL G+ + H + H R FAT+ ++L
50  Sbjct: 127  ADNRGFDVTLVRDATATHDR-TLDERLPPSVVHRTALAHLRGEFATLATTATVL 180

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5610 (GBS652) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 133 (lane 2 & 3; MW 49.7kDa) + lane 4; MW 27kDa) and in Figure 186 (lane 9; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell

-2040-

extract is shown in Figure 133 (lane 5 & 7; MW 24.8kDa) and in Figure 178 (lane 10; MW 25kDa). Purified GBS652-GST is shown in Figure 243, lane 9; purified GBS652-His is shown in Figure 229, lane 10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1804

A DNA sequence (GBSx1911) was identified in *S.agalactiae* <SEQ ID 5611> which encodes the amino acid sequence <SEQ ID 5612>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0945(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14504 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
Identities = 55/118 (46%), Positives = 82/118 (68%)
```

```
Query: 1  MTEKDLLQLVVKAADEKRAEDIVILDLPVTSVADYFVIMSASNSRQLEAIADNIREQVK 60
      M +K +L++  A D+KRAEDI+ LD++ ++ VADYF+I  ++ +Q++AIA I++Q
Sbjct: 1  MNQKSILKIAAAACDDKRAEDILALDMEGISLVADYFLICHGNSDKQVQAIAREIKDQAD 60

Query: 61  GNGGDASHLEGDSKAGVLLDLNSVVVHIFSEDERQHYNLEKLWHEAPLLDAEVFMTE 118
      NG  +EG +A WVL+DL VVVH+F +DER +YNLEKLW +APL D + M +
Sbjct: 61  ENGIQVKKMEGFDEARWVLVDLGDVVVHVHFKDERSYNNLEKLWGDAPLADLDFGMNQ 118
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5613> which encodes the amino acid sequence <SEQ ID 5614>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.69    Transmembrane    91 - 107 ( 91 - 107)

----- Final Results -----
      bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14504 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
Identities = 55/113 (48%), Positives = 80/113 (70%)
```

```
Query: 17  MKKEELLKIVVEATEEKRAKDILALDLEGLTSLTDYFVIASATNSRQLEAIADNIREKVK 76
      M ++ +LKI  A ++KRA+DILALD+EG++ + DYF+I  + +Q++AIA I+++
Sbjct: 1  MNQKSILKIAAAACDDKRAEDILALDMEGISLVADYFLICHGNSDKQVQAIAREIKDQAD 60

Query: 77  EAGGDASHVEGNSQAGWLLDLTDVVVHFLFEDERYHYNLEKLWHEAPALD 129
      E G  +EG +A WVL+DL DVVVH+F +DER +YNLEKLW +AP LD
Sbjct: 61  ENGIQVKKMEGFDEARWVLVDLGDVVVHVHFKDERSYNNLEKLWGDAPLADLD 113
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 78/116 (67%), Positives = 100/116 (85%)
```

```
Query: 1  MTEKDLLQLVVKAADEKRAEDIVILDLPVTSVADYFVIMSASNSRQLEAIADNIREQVK 60
```


-2041-

```

M +++LL++VV+A +EKRA+DI+ LDL+ +TS+ DYFVI SA+NSRQLEAIADNIRE+VK
Sbjct: 17 MKKEELLKIIVVEATEEKRAKDILALDLEGLTSLTDYFVIASATNSRQLEAIADNIREKVK 76

Query: 61 GNGGDASHLEGDSKAGWVLLDLNSVVVHIFSEDERQHYNLEKLWHEAPLLDAEVFM 116
          GGDASH+EG+S+AGWVLLDL VVVH+F EDER HYNLEKLWHEAP + + ++
Sbjct: 77 EAGGDASHVEGNSQAGWVLLDLTDVVVHLEFEDERYHYNLEKLWHEAPVALDAYL 132

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1805

A DNA sequence (GBSx1912) was identified in *S.agalactiae* <SEQ ID 5615> which encodes the amino acid sequence <SEQ ID 5616>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2415(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 1806

A DNA sequence (GBSx1913) was identified in *S.agalactiae* <SEQ ID 5617> which encodes the amino acid sequence <SEQ ID 5618>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1570(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB14503 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
Identities = 86/242 (35%), Positives = 154/242 (63%), Gaps = 4/242 (1%)

```

```

Query: 4 YETFAAVYDAVMDDTLYAKWTFSLRHFPGKKKLELACGTGIQSVRFQAQYAVTGID 63
Y+ FA+VYD +M Y +WT + P+ K ++L+LACGTG S+R A+ G+ VTG+D
Sbjct: 3 YQGFASVYDELMASHAPYDQWTKWIEASLPE-KGRILDACGTGEISIRLAEGFVETGID 61

```

```

Query: 64 LSGDMLKLAKKRATSAHQSTQFIEGNMLDLSNV-GKYDLITCYSDSICYMQDEVEVGDFV 122
LS +ML A+++ +S+ Q I F++ +M +++ G++D + DS+ Y++ + +V + F
Sbjct: 62 LSEEMLSFAQQKVSSS-QPILFLQDMREITGFDGQFDAVVICCDLSNLYLKTNDVIETF 120

```

```

Query: 123 IEVYKALEENGVFIFDVHSTYQTDKVFPGYSYHENADDFAMVWDTYEDDAPHSIVHELTF 182
V++ L+ G+ +FDVHS+++ +VFP ++ + +D + +W ++ S++H+++F
Sbjct: 121 KSVFRVLKPEGILLFDVHSSFKIAEVFPDSTFADQDEDISYIWQSFAGSDELSVIHDMSE 180

```

```

Query: 183 FVQEEEDGRFTRHDEVHEERTYDILTYDILLEQAGFKDKVKYADFEDKKPTATSARWFFVA 242
FV + + R DE HE+RT+ + Y+ +L+ GF+ +V ADF D +P+A S R FF A
Sbjct: 181 FVWNGEA-YDRFDETHEQRTFFPVEEYEEMLKNCGFQLHRVTADFDTPEPSAQSERLFFKA 239

```

-2042-

Query: 243 HK 244
K
Sbjct: 240 QK 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5619> which encodes the amino acid sequence <SEQ ID 5620>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2315(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/243 (78%), Positives = 215/243 (87%), Gaps = 2/243 (0%)

Query: 4 YETFAAVYDAVMDDTLyakWDFSLRHFPK--GKKKLELACGTGIQSVRFAQAGYAVTG 61
YE FA+VYDAVMDD+LY WDFSLRH PK G+ +LELACGTGIQSVRFAQAG+ VTG
Sbjct: 21 YEKFASVYDAVMDDSLYDLWDFSLRHLPKSKGRNRLLELACGTGIQSVRFAQAGFDVTG 80

Query: 62 LDLSGDMKLAKKRATSAHQSIQFIEGNMLDLNsvgKYDLITCYSDSICYMQDEVEVGDV 121
LDLS DML +AKKRA SA + I FI+GNMLDLN VG++D +TCYSDSICYMQDEV+VGDV
Sbjct: 81 LDLSQDMLAIAKKRAQSAKKKIDFIQGNMLDLNsvGVQFDFVTCTCYSDSICYMQDEVVDVGDV 140

Query: 122 FIEVYKALEENGVFIFDVHSTYQTDKVFPGYSYHENADDFAMVWDITYEDDAPHsIVHELT 181
F EVY L +G+FIFDVHSTYQTD+ FPGYSYHENADDFAMVWDITY D+APHS+VHELT
Sbjct: 141 FKEVYDVLANDGIFIFDVHSTYQTDCECFPGYSYHENADDFAMVWDITYADEAPHSVVHELT 200

Query: 182 FFVQEEEDGRFTRHDEVHEERTYDILTYDILLEQAGFKDVKVYADFEDKKPTATSARWFFV 241
FF+QE+DGRF+R DEVHEERTY++LTYDILLEQAGFK KVVYADFEDK+PT TS RWFFV
Sbjct: 201 FFIQEDDGRFSRFDEVHEERTYELLTYDILLEQAGFKSFKVYADFEDKEPTKTSKRWFFV 260

Query: 242 AHK 244
A+K
Sbjct: 261 AYK 263

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1807

A DNA sequence (GBSx1914) was identified in *S.galactiae* <SEQ ID 5621> which encodes the amino acid sequence <SEQ ID 5622>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3538(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06304 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 129/367 (35%), Positives = 184/367 (49%), Gaps = 45/367 (12%)

Query: 1 MIVTGIVAEPNPFHNGHKYLLEQAQ-----GIKVIAMSGNFMQGEPAIVDKWTRSQMAL 55
M G+V E+NPFHNGH + L +A+ + + MSG F+QGEPAI+ KW R+ +AL
Sbjct: 1 MKAVGVVVEYNPFHNGHLHLTEARKQAKADVVIAMSGYFLQGEPAILPKWERTSLAL 60

-2043-

Query: 56 ENGADLVIELPFLVSVQSADYFASGAVSILARLGVDNLFCFGTEE--MLDYARIGDIYVNK 113
 + GADLV+ELP+ S Q A++FA+GAVSILA L D LCFG+EE + + R+
 Sbjct: 61 QGGADLVVELFPYAFSTQKAWEFATGAVSILAALADALCFGSEEGTIEPFHRLYHFMMAKH 120

5 Query: 114 KEEMEAFLLKKQSD-SLSYPQKMQAMWQEFAGIT--FSGQTPNHILGLAYTKAA--SQNGI 168
 + + +K++ D +SYP ++ G PN+ILG Y KA I
 Sbjct: 121 RLAWDRMIKEELDKGMSYPTATSLAFKRLEGSAEHLDLSPNNILGFHYVKAIYDLHTSI 180

10 Query: 169 RLNPFIQRQAGYHSSEKTE-IFASATSLRK-----HQSDRFF-----VEKGMPNSD 213
 + I R AGYH E ASATS+RK DR + K
 Sbjct: 181 KAMTIPRIKAGYHDDSLNESSIASATSIRKSLKTKEGWQMVDRVVPSTTEMLKSFEKET 240

15 Query: 214 LFLNSPQVWQDYFSLKYLKQIMTHS--DLTQIYQVNEEIANRIKSQIRYVETVDELVDKV 271
 FL S W+ F LLKY+++T + L IY+ E + R I + + + K+
 Sbjct: 241 TFLPS---WERLFPLLKYRLLTATPEQLHAIYEGEEGLETRALKTIVSATSFHDWMTKM 296

20 Query: 272 ATKRYTKARIRRLTYILINAVESPIPN-----IHVLGFTQKGQHLKSVKK-- 319
 TKRYT RI+R T++ N + I + I +LG T +GQ +L KK
 Sbjct: 297 KTKRYTWTRIQRYATHLFTNTTKEEIHVSFLPRGTESLPYIRLLGMTSRGQMYLNGKKKQL 356

Query: 320 SVDIVTR 326
 + ++TR
 Sbjct: 357 TTPVITR 363

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5623> which encodes the amino acid sequence <SEQ ID 5624>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3165(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 221/359 (61%), Positives = 288/359 (79%)

Query: 1 MTVTGTGIAEFNPFHNGHKYLLEQAQGIKVIAMSGNFMQRGEPALVDKWTRSQMALENGAD 60
 MTVTGTGIAEFNPFHNGHKYLLE A+G+K+IAMSGNFMQRGEPAA++DKW RS+MAL+NGAD
 40 Sbjct: 1 MTVTGTGIAEFNPFHNGHKYLLETAEGLKIIAMSGNFMQRGEPALIDKWIRSEMALKNGAD 60

Query: 61 LVIELPFLVSVQSADYFASGAVSILARLGVDNLFCFGTEEMLDYARIGDIYVNKKEEMEAF 120
 +V+ELPF VSVQSADYFA GA+ IL +LG+ L FGTE ++DY ++ +Y K E+M A+
 45 Sbjct: 61 IVVELPFFVSVQSADYFAQGAIDILCQLGIQQLAFGTENVIDYQKLIKVEKKSEQMTAY 120

Query: 121 LKKQSDLSYPQKMQAMWQEFAGITFSGQTPNHILGLAYTKAASQNGIRLNPIQRQAGY 180
 L D+ SYPQK Q MW+ FAG+ FSGQTPNHILGL+Y KA++ I+L PI+RQGA Y
 50 Sbjct: 121 LSTLEDTFSTPQKTQKMWEIFAGVKFSGQTPNHILGLSYAKASAGKHIQLCPIKRGGAAY 180

Query: 181 HSSEKTEIFASATSLRKHQSDRFFVEKGMPNSDLFLNSPQVWQDYFSLKYLKQIMTHSDL 240
 HS +K + ASA+++R+H +D F+ +PN+ L +N+P + W YFS LKYQI+ HSDL
 55 Sbjct: 181 HSKDKNHLLASASAIRQHLNDWDFISHSVPNAGLLINPHMSWDHYFSFLKYQILNHSDL 240

Query: 241 TQIYQVNEEIANRIKSQIRYVETVDELVDKVATKRYTKARIRRLTYILINAVESPIPN 300
 T I+QVN+E+A+RIK I+ + +D LVD VATKRYTKAR+RR+LTYIL+NA E +P
 60 Sbjct: 241 TSIFQVNDELASRIKKAIVKSNIDHLVDTVATKRYTKARVRRILTYILVNAKEPTLPKG 300

Query: 301 IHVLGFTQKGQHLKSVKKSVDIVTRIGSQTWDSLTQRADSVYQMGNANIAEQTWGRIP 359
 IH+LGFT KGQ HLK +KKS ++TRIG++TWD +TQ+ADS+YQ+G+ +I EQ++GRIP
 60 Sbjct: 301 IHILGFTSKGQAHLKKLKKSRPLITRIGAETWDEMTQKADSIYQLGHQDIPEQSFGRIIP 359

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2044-

Example 1808

A DNA sequence (GBSx1915) was identified in *S.agalactiae* <SEQ ID 5625> which encodes the amino acid sequence <SEQ ID 5626>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```

5   Possible site: 17
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3117(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
    Identities = 91/246 (36%), Positives = 144/246 (57%), Gaps = 14/246 (5%)

    Query: 4   VKEISHISGISVRTLHYYDEIDLLSPSFVGENGYRYYDDESLIKLQEILLFKELEFPLKK 63
              VK+++ ISG+S+RTLH+YD I+LL+PS + + GYR Y D L +LQ+IL FKE+ F L +
    Sbjct: 5   VKQVAEISGVSIRTLLHHYDNIELLNPSALTADAGYRLYSADLERLQILFFKEIGFRLDE 64

20   Query: 64  IKEIMDSPNYDRNQALLDQIRWLELKKQRLEEVIEHAK----SIQRGKNMSD---FTAYN 116
              IKE++D PN+DR AL Q L KKQR++E+I+ S+ G+ M+ F +
    Sbjct: 65  IKEMLDHPNFRKAALQSQKEILMKKKQRMDEMIQTIDRTLLSVDGGETMKNKRDLFAGLS 124

25   Query: 117 QEELEAFQ----EEARTRWGD--TDSYKEFENSHSKNDFSMISQAMSQIFKDFGQLKELS 170
              +++E Q +E R +G + ++ +++S +D+ I I++ +
    Sbjct: 125 MKDIEEHQQTFADEVKLYGKEIAEETEKRTSAYSADDWRTIMAEFDSIYRRIAARMKHG 184

30   Query: 171 PTDEKVQKQVQILQDYITAQFYNCNTNDLLASLGIMYIQDERFQKSIDNWGGQGTALFVSK 230
              P D ++Q V +D+I Y+CT D+ LG +YI DERF SI+ + G+G A F+ +
    Sbjct: 185 PDDAEIQAAVGAFRDHICQYHYDCTLDIFRGLGEVYITDERFTDSINQY-GEGLAFLRE 243

    Query: 231 AIDSYC 236
              AI YC
35   Sbjct: 244 AIIYYC 249

```

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1809

A DNA sequence (GBSx1916) was identified in *S.agalactiae* <SEQ ID 5627> which encodes the amino acid sequence <SEQ ID 5628>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 39
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

55   >GP:CAB14597 GB:Z99117 yrkC [Bacillus subtilis]
    Identities = 56/129 (43%), Positives = 74/129 (56%), Gaps = 7/129 (5%)

    Query: 2   KGFHGNIEKLTGNTNFRQVLYTAEHCQLVLMTLFVGGGEIGSEIHAENDQFFRFEGHKG 61
              K F NI + T N FR L+T +H Q+ LM+L +G +IG EIH DQF R E G G

```

-2045-

Sbjct: 59 KPFVVNINRATKQNTFR TALWTGKH FQVTLMSLGIGEDIGLEIHPNVDQFLRIEQGRGI 118

Query: 62 VVIDGN-----EYEVADGD AII VPAGAEHNVINTSETEMLKLYTIYSPAHHKDGII RAT 115
 V + + + V D AI+VPAG HNVINT T LKLY+IY+P +H G + T

5 Sbjct: 119 VKMGKSKDHLNFQRNVYDDSAIVVPAGTWHNVINTGNTF-LKLYSIYAPPNHPFGTVHET 177

Query: 116 REEAENEENE 124
 + +A E+

10 Sbjct: 178 KADAVAAED 186

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1810

15 A DNA sequence (GBSx1917) was identified in *S.agalactiae* <SEQ ID 5629> which encodes the amino acid sequence <SEQ ID 5630>. This protein is predicted to be glycerol uptake facilitator (glpF). Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have an uncleavable N-term signal seq

20 INTEGRAL Likelihood = -9.08 Transmembrane 156 - 172 (153 - 180)
 INTEGRAL Likelihood = -6.21 Transmembrane 135 - 151 (132 - 155)
 INTEGRAL Likelihood = -4.09 Transmembrane 86 - 102 (80 - 103)
 INTEGRAL Likelihood = -3.93 Transmembrane 213 - 229 (212 - 230)
 25 INTEGRAL Likelihood = -3.72 Transmembrane 8 - 24 (5 - 28)
 INTEGRAL Likelihood = -2.76 Transmembrane 38 - 54 (36 - 58)

----- Final Results -----

30 bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04811 GB:AP001510 glycerol uptake facilitator [Bacillus halodurans]
 Identities = 135/230 (58%), Positives = 171/230 (73%)

35 Query: 1 MTQFLGEFLGT FILVLLGDGVVAGNVLSKTKEEGTGWTAIVFGWGIAC TVAVVVSGLFSP 60
 M+ FLGE +GT IL++LG GVVAG VL TK E GW I WG+A AVY G S
 Sbjct: 1 MSPPLGEVIGT MIIILGGGVVAGVVLKGT KSENGGWIVITA AWGLAVATAVYCVGQISG 60

40 Query: 61 AHLNPAVTLAMASIG AISWGQVIPFIIAQMLGAMVAATILWLHYYPHWKETKDSGLILAS 120
 AHLNPAVT+ +A +GA W QV +I+AQMLGAM+ AT+++LHYYPH+K T+D G LA
 Sbjct: 61 AHLNPAVTIGLALVGAFWSQVAGYIVAQMLGAMIGATLVFLHYYPHFKATEDQGAKLAV 120

45 Query: 121 FSTGPAIRHTPSNLLGEIIGTAILVITIMAI GPSKVAAGLGPIIVGIVIFAVGFSLDPTT 180
 FST PAI+H P+N E++GT +LV+ I+AI G ++ GL P+IVG++I +G SL TT
 Sbjct: 121 FSTDPAIKHLPANFFSEVLGT FVLVLGILAIGANEFT EGLNPLIVGLLIVVIGLSLGTT 180

50 Query: 181 GYAINPARDLGPRIMHAILPIENKCNDSWYAWIPVVGPIIGGVLGAILY 230
 GYAINPARDLGPR+ H +LPI KG+S+WSYAWIP+VGPIIGG +GA+ Y
 Sbjct: 181 GYAINPARDLGPRIAHFLLP IPGKSSNWSYAWIPVVGPIIGGGIGALTY 230

There is also homology to SEQ ID 2854.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2046-

Example 1811

A DNA sequence (GBSx1918) was identified in *S.agalactiae* <SEQ ID 5631> which encodes the amino acid sequence <SEQ ID 5632>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1694 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07114 GB:AP001518 unknown conserved protein in others

[Bacillus halodurans]

Identities = 64/118 (54%), Positives = 85/118 (71%)

Query: 5 GIIVVSHSKNIAQGVVDLISEVAKDVSITYVGGTDEGEIGTSFQVQQIVEQNDKKTLLA 64
GI++ SH +A+G+V L+ E AKDVSITY GGT+D ++G SF+++QQ V N+ L

Sbjct: 7 GIVISSHPALAEIGIVTLLKEAAKDVSTYAGGTDDQVGASFQIQAVMDNEADELFV 66

Query: 65 FFDLGSAKMNLVADFSEKNIIINSVPVVEGAYTAAALLOAGADLDSIQSQLAELTI 122
F+DLGSAKMN+E+V + SEK I + V +VEGAYTAAAL Q GA ++I QL LTI

Sbjct: 67 FYDLGSAKMNVEMVMESEKTIHLMDVALVEGAYTAAALTQGGASFETIMQQLPLTI 124

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1812

A DNA sequence (GBSx1919) was identified in *S.agalactiae* <SEQ ID 5633> which encodes the amino acid sequence <SEQ ID 5634>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4753 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07115 GB:AP001518 unknown conserved protein in others

[Bacillus halodurans]

Identities = 98/190 (51%), Positives = 135/190 (70%), Gaps = 2/190 (1%)

Query: 3 VKTAIEWMHTFNQKIQSNDYLSLDTPIGDGDHGNMARGMTAVIENLDNNEFSSAADV 62
V+ +W+H F++K+Q+N+ YLSELD+ IGDGDHG NMARG+ V L N F S +V

Sbjct: 4 VENTTKWLHAFHEKVQANQSYLSLDSAIGDGDHGTNMARGLAEVERKLKENLFESPQEV 63

Query: 63 FKTVSMQLLSKVGASGPLYGSAFMGITK-AEQSKSTISEALGAGLEMIQKRGKAELENEK 121
K +M L+SK GGASGPLYG+A + ++K I +++ AGL I KRGKA EK

Sbjct: 64 LKMAAMALISKTTGGASGPLYGTALLEMSKQVANDPQNIGKSIEAGLNGILKRGKATTGEK 123

Query: 122 TMVDVWHGVIEAI-EKNELTEDRIDSLVDATKGMKATKGRASYVGERSVGHIDPGSFSSG 180
TMVD+W V+E++ + +L+++RI V TK MKATKGRASY+GERS+GH+DPG+ SSG

Sbjct: 124 TMVDIWKPVVESLMABQQLSKERIQQFVSETKEMKATKGRASYLGERSLGHLDPGAVSSG 183

Query: 181 LLFKALLEVG 190

LF+A+++ G

-2047-

Sbjct: 184 YLFEAMIDGG 193

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1813

A DNA sequence (GBSx1920) was identified in *S.agalactiae* <SEQ ID 5637> which encodes the amino acid sequence <SEQ ID 5638>. This protein is predicted to be dihydroxyacetone kinase (b1200). Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2080(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA07116 GB:AP001518 dihydroxyacetone kinase [Bacillus halodurans]
Identities = 204/329 (62%), Positives = 261/329 (79%)

Query: 1  MKKILNQPTDVVTEMLDGLAYVHNDLVHRIEGFDIIARNEEKSGKVALISGGGSGHEPSH 60
          MKKIILN P +V+ EMLDG Y + LV R+ G +I R E GKVAL+SGGGSGHEPSH
Sbjct: 1  MKKIILNDPQNVLD EMLDGFVYANGHLVERVAGTGVIRRTYEDKGKVALVSGGGSGHEPSH 60

Query: 61  AGFVGEGMLSAAVCGAVFTSPTPDQVLEAIKEADGAGVFMVKNYSGDIMNFEMAQDMA 120
          AGFVG+GMLSAAVCG VFTSPTPDQ+ E IK AD+G GV ++IKNY+GD+MNFEMA +MA
Sbjct: 61  AGFVGQGMLSAAVCGEVFTSPTPDQIFEGIKAADQGGGVLLIIKNYTGDMNFMEMAGEMA 120

Query: 121  EMEGIEVASVVDDDDIAVEDSLYTGQKRGVAGTILVHKILGHAARHGKSLQEIKAIADDEL 180
          E EGI V ++V+DDIAVEDS +T G+RGVAGTI+VHKI+G AA G SLQ +K + + +
Sbjct: 121  EAEGITVDHIIVNDDIAVEDSSFTAGRRRGVAGTIIIVHKIVGAAAEAGLSLQSLKVLGETV 180

Query: 181  VPNIHTVGLALSGATVPEVGKPGFVLAEDEIEFGIGIHGEPGYRKEKMQPSKALATELVD 240
          + N T+G+++ ATVP VGKPGF L +DE+E+G+GIHGEFGYRKEK++ SK +A EL+
Sbjct: 181  IENTKTIGVSILPATVPAVGKPGFELGDDEMEYGVGIGHGEPGYRKEKLSKEIAEELIL 240

Query: 241  KLIESFDAKSGEKYGVLLINGMGATPLMEQYVVFANDVAKLLEDKGIEVNYKKLGNYMTSID 300
          KL E+F G+KYGV L+NG+GATPLMEQYVF NDVA L ++G+ + +KK+G++MTSID
Sbjct: 241  KLKEAFGWSKGDKYGVLVNGLGATPLMEQYVFMNDVANKLTEEGLNIQFKKVGSEFMTSID 300

Query: 301  MAGLSLTLIKLENQEWLEALNSDVTITIAW 329
          MAG+SLTLIK+ ++WL+ N +V T+ W
Sbjct: 301  MAGVSLTLIKIVEERWLDYWNHEVKTVDW 329

```

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1814

50 A DNA sequence (GBSx1921) was identified in *S.agalactiae* <SEQ ID 5639> which encodes the amino acid sequence <SEQ ID 5640>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1997(Affirmative) < succ>

```

-2048-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BA07113 GB:AP001518 unknown [Bacillus halodurans]
 Identities = 59/142 (41%), Positives = 82/142 (57%), Gaps = 5/142 (3%)

Query: 1 MTSSLITKKKIAKSFKRLFISQAFDKISVSDIMEDAGIRRTFYNHFDKYALLEWIFQT 60
 MT+S+ITKK IAK+FK L Q F KISVSDIM A +RRQTFY HF DK+ LL WI++
 10 Sbjct: 1 MTNSIITKKVIAKAFKDLMEVQPFKISVSDIMNRANMRRQTFYHFQDKFELLHWIYKQ 60

Query: 61 ELSEQVTDNLDYISGFQQLSELLTFFKMNQEFYIKLFQIEDQNDFFSSYFESYCEQLVDKL 120
 E E D L Y + L+ +F NQ FY + + QN F+ Y + + L
 Sbjct: 61 ETKEHSIDFLAYDDIHTIFRHLMHYFYENQTFYQRAMVVNGQNGFTDYLVEHIQTL---Y 117

15 Query: 121 LSDYSKSNFNQKERVTFINYHS 142
 L++ + +QK+R +++S
 Sbjct: 118 LNEIDRR--SQKREFISSFYS 137

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5641> which encodes the amino acid sequence <SEQ ID 5642>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2101(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 31/115 (26%), Positives = 58/115 (49%), Gaps = 6/115 (5%)

Query: 7 TKKKIAKSFKRLFISQAFDKISVSDIMEDAGIRRTFYNHFDKYALLEWIFQTELSEQV 66
 TK + + L Q+F+ ++VSD+ + AGI R TFY H+ DK+ ++ F+ + + +
 35 Sbjct: 8 TKAYVK TALTTLLTEQSFETLTVDLTKKAGINRGTFYLYHTDKFDMNH-FKNDTLDDL 66

Query: 67 TDNLD----YISGFQQLSELLTFFKMNQEFYIKLFQIEDQNDFFSSYFESYCEQLV 117
 L+ Y Q+L++ L++ ++EF L I F + +C Q +
 40 Sbjct: 67 YRLINQAEIYTDTRQVLNQTLSYLIEHREFITALATI-SYLFQPLIKDFCYQFL 120

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1815

45 A DNA sequence (GBSx1922) was identified in *S.agalactiae* <SEQ ID 5643> which encodes the amino acid sequence <SEQ ID 5644>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1974(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-2049-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1816

A DNA sequence (GBSx1923) was identified in *S.agalactiae* <SEQ ID 5645> which encodes the amino acid sequence <SEQ ID 5646>. This protein is predicted to be dihydroxyacetone kinase (b1200). Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1806(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07112 GB:AP001518 dihydroxyacetone kinase [Bacillus halodurans]
 Identities = 141/285 (49%), Positives = 197/285 (68%), Gaps = 1/285 (0%)

20 Query: 45 IPILSGGGSGHEPAHFGYVGEGLSAAISGPIFVPPCASDILETIRFINRGKGVFVIKN 104
 +PI+SGGGSGHEP H GYVGEGL+AA+ G +FVPP A +L IR +++GKGV +IKN
 Sbjct: 46 VPIISGGSGGHEPGHLGYVGEGLAAAVHGDVFPVPSAQVLAIRQMDQKGVLLIKN 105

Query: 105 FEADLEEFSAIEQARQEGIPKIYIVSHDDISVET-SNFKIRHRGVAGTVLLHKKIIGQAA 163
 F ADL F A QAR EG + +++ +DD+SVE+ ++F+ R RGVAG VL+HKIIG AA
 25 Sbjct: 106 FVADLATFLSAEVQARAEGRDVAHVIVNDDVSVESDASFEKRRRGVAGAVLVHKIIGAAA 165

Query: 164 LEGASLDELEQLGLSLTSMATLGVASKSATILGQHQPVFIDIEGYISFGIGIHGEPGYR 223
 EG SL+ L+++G + ++ATLGVA A + + +P F +EEG + FG+GIHGE GYR
 30 Sbjct: 166 KEGYSLEALQEIGEQQVKNLATLGVALTHADLPERREPQFLLEEGEVYFGVGIHGEQGYR 225

Query: 224 TMPFVSMEHLANELVNLKMKLRWQDGEAFILLINNLGGSSKMEELLFTNAVMEFLALDD 283
 VS E LA ELVNKLK RW + + +LIN LGG+ +E+ +F N V LA+++
 35 Sbjct: 226 KEKLVSSSELLAVELVNKLKSLYRWDKNDQYAVLINGLGGTPLIEQYVFANDVRLLAIEN 285

Query: 284 LQLPFIKTGHLITSLDMAGLSVTLCRVKDSRWIDYLLKHKT DARAW 328
 L + F+K G +TSL+M G+S+T+ ++ D +W+ +L D W
 Sbjct: 286 LHVSVVKVGTQLTSLNMKGISLTMLKICDEQVWKWLYAPVDVAHW 330

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1817

A DNA sequence (GBSx1924) was identified in *S.agalactiae* <SEQ ID 5647> which encodes the amino acid sequence <SEQ ID 5648>. Analysis of this protein sequence reveals the following:

45 Possible site: 53

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3902(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10085> which encodes amino acid sequence <SEQ ID 10086> was also identified.

-2050-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC75047 GB:AE000290 orf, hypothetical protein [Escherichia coli K12]
Identities = 182/237 (76%), Positives = 201/237 (84%)

5 Query: 20 MGRKWANIVAKKTAKDGANSKVYAKFGVEIYVAAKQGEPPESNSALKFVLDRAKQAQVP 79
MGRKWANIVAKKTAKDGA SK+YAKFGVEIY AAKQGEPPDPE N++LKFV++RAKQAQVP
Sbjct: 1 MGRKWANIVAKKTAKDGATSKIYAKFGVEIYAAKQGEPPDPELNTSLKFVIERAKQAQVP 60

10 Query: 80 KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 139
KHVIDKAIDKAKG DETFV+GRYEGFGPNGSMII +TLTSNVNRT ANVRT + K GGN
Sbjct: 61 KHVIDKAIDKAKGGDETFVQGRYEGFGPNGSMIIAETLTSNVNRTIANVRTIFNKKGGN 120

15 Query: 140 MGASGSVSYLFDKKGVI VFAGDDADTVFEQLLEADVDDVEAEEGTITVYTAPTDLHKG 199
+GA+GSVSY+FD GVIVF G D D +FE LLEA+VDV DV EEG I +YT PTDLHKG
Sbjct: 121 IGAAGSVSYMFDNTGVIVFKGTDPDHIFEILLEAEVDVRDVTEEGNIVIYTEPTDLHKG 180

20 Query: 200 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVA 256
I AL+ G+ EF TELEMI QSEV L +DLE FE L+DALE DDDVQKVYHNVA+
Sbjct: 181 IAALKAAGITEFSTTELEMIQSEVELSPEDLEIFEGLVDALEDDDDVQKVYHNVA 237

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5649> which encodes the amino acid sequence <SEQ ID 5650>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2926(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 233/238 (97%), Positives = 236/238 (98%)

35 Query: 20 MGRKWANIVAKKTAKDGANSKVYAKFGVEIYVAAKQGEPPESNSALKFVLDRAKQAQVP 79
MGRKWANIVAKKTAKDGA SKVYAKFGVEIYVAAKQGEPPDPE N+ALKFV+DRAKQAQVP
Sbjct: 1 MGRKWANIVAKKTAKDGATSKVYAKFGVEIYVAAKQGEPPDPELTALKFVIDRAKQAQVP 60

40 Query: 80 KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 139
KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN
Sbjct: 61 KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 120

45 Query: 140 MGASGSVSYLFDKKGVI VFAGDDADTVFEQLLEADVDDVEAEEGTITVYTAPTDLHKG 199
MGASGSVSYLFDKKGVI VFAGDDAD+VFEQLLEADVDDVEAEEGTITVYTAPTDLHKG
Sbjct: 121 MGASGSVSYLFDKKGVI VFAGDDADSVFEQLLEADVDDVEAEEGTITVYTAPTDLHKG 180

Query: 200 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVA 257
IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVA 257
Sbjct: 181 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVA 238

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1818

A DNA sequence (GBSx1925) was identified in *S.agalactiae* <SEQ ID 5651> which encodes the amino acid sequence <SEQ ID 5652>. Analysis of this protein sequence reveals the following:

55 possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2507(Affirmative) < succ>

-2051-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1819

A DNA sequence (GBSx1926) was identified in *S.agalactiae* <SEQ ID 5653> which encodes the amino
10 acid sequence <SEQ ID 5654>. Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1523 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAA20826 GB:AL031541 hypothetical protein SCI35.37 [Streptomyces
coelicolor A3(2)]
Identities = 73/178 (41%), Positives = 101/178 (56%), Gaps = 2/178 (1%)
Query: 35 VKNAGGLPVILPISEAESAKAYVEMIDKLIISGGQNVLPSSYYGEEKIIESDDYSLARDIF 94
25 V+ AGGL +LP E A A V +D ++I+GG +V P YG E + + ARD +
Sbjct: 37 VQRAGGLAAMLPPDAPEHAATVARVDGVVIAGGPDVEPVRYGAEPDPRGTGPPARARDTW 96
Query: 95 EFALVEEALKQNKPIFAICRGMQLVNVALGGTLNQSIDNHQEPYIGFAHYLNVEKGSFL 154
30 E AL+E AL P+ ICRGMQL+NVALGGTL Q I+ H + + H + G+
Sbjct: 97 ELALIEAALAARVPLLIGICRGMQLLNVALGGTLVQHIERHAEVVGVFGGHPVRPVPPTLY 156
Query: 155 EGFISGDFKINSLHRQSVKLLAEGILVSARDPRDGTVEAYESRT-EQCTIGVQWHPPEL 211
G + + + + H Q+V L GL+ SA DGTVEA E + ++GVQWHPPE+
35 Sbjct: 157 AGAVPEETFVPTYHHQAVDRLGSGLVASAH-AADGTVEALEMPGSGGWVLGVQWHPPEM 213

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5655> which encodes the amino acid
sequence <SEQ ID 5656>. Analysis of this protein sequence reveals the following:

40 Possible site: 52
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.1210 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/230 (48%), Positives = 145/230 (62%), Gaps = 3/230 (1%)
Query: 2 LTKPIIGITGNEREMSDIPGYYSVSRHISEGVKNAGGLPVILPISEAESAKAYVEMID 61
50 +TKPIIGIT N+R + + + V +GGLP++LPI + +AK YV M+D
Sbjct: 1 MTKPIIGITANQRLNMLDNLFWSYAPTGFVQAVTQSGGLPLLLPIGDEAAAKTYVSMVD 60
Query: 62 KLIISGGQNVLPSSYYGEEKIIESDDYSLARDIFEALVEEALKQNKPIFAICRGMQLNVN 121
K+I+ GGQNV P YY EEK DD+S RD FE A+++EA+ KPI ICRG QL+NV
55 Sbjct: 61 KIILIGGQNVDPKYYQEKAADFDDFSPERDTFELAIKEAITLKKPILGICRGTQLMNV 120

-2052-

Query: 122 ALGGTLNQSIDNHQYQ-PYIGFAHYLNVEKGSFLEGFISGDFKINSLHRQSVKLLAEGLI 180
 ALGG LNQ ID+H+QE P +H + +E S L INS HRQS+K +A+ L
 Sbjct: 121 ALGGNLNQHIDSHWQEAPSDFLSHEMIIEPDSILYPTIYGHKTINSFHRQSLKTVAKDLK 180

5 Query: 181 VSARDPRDGTVEAYESRTEQC-IIGVQWHPMLH-QIENQTLFGYFVNE 228
 V ARDPRDGT+EA S + +GVQWHPEL+ + E+ LF FVN+
 Sbjct: 181 VIARDPRDGTIEAVISTNDAIPFLGVQWHPMLQGVRDEDLQLFRLFVND 230

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1820

A DNA sequence (GBSx1927) was identified in *S.agalactiae* <SEQ ID 5657> which encodes the amino acid sequence <SEQ ID 5658>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5794(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1821

A DNA sequence (GBSx1928) was identified in *S.agalactiae* <SEQ ID 5659> which encodes the amino acid sequence <SEQ ID 5660>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0524(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8905> which encodes amino acid sequence <SEQ ID 8906> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 22 Crend: 4
 McG: Discrim Score: 8.37
 GvH: Signal Score (-7.5): -0.64
 Possible site: 21
 >>> May be a lipoprotein
 ALOM program count: 0 value: 6.74 threshold: 0.0
 PERIPHERAL Likelihood = 6.74 112
 modified ALOM score: -1.85

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

-2053-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2919> which encodes the amino acid sequence <SEQ ID 2920>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 120/162 (74%), Positives = 141/162 (86%), Gaps = 5/162 (3%)

Query: 6 LAACSSKSHTTKTGK----KEVNFATVGTTAPFSYVKDGLTGFDIEVAKAVFKGSDNYK 61

LAAC S S T ++G KEV FATVGTTAPFSY K G+LTG+DIEVAKAVFKGSD+YK

Sbjct: 20 LAACGS-SKTAESGNQGSSEVLFATVGTTAPFSYEKGGQLTGYDIEVAKAVFKGSDDYK 78

Query: 62 VTFKKTEWSSVFTGIDSGKFQMGNNISYSSERSQKYLFSYPIGSTPSVLAVPKNSNIKA 121

V+FKKTEWSS+FTG+DSGK+QMGNNIS++ ERS KYLFSYPIGSTPSVL VPK+S+IK+

Sbjct: 79 VSFKKTEWSSIIFTGLDSGKYQMGNNISFTKERSAKYLFSYPIGSTPSVLVVPKDSDIKS 138

Query: 122 YNDISGHKTQVVQGTITAKQLENFNKEHQKNPVTCLKYTNNEL 163

++DI GH TQVVQGTIT+ QLE+FNK+H NPVTLK+TNEN+

Sbjct: 139 FDDIQGHITTQVVQGTTSVAQLEDFNKKHSDNPVTCLKFTNENI 180

SEQ ID 8906 (GBS71) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 4; MW 31.8kDa).

GBS71-His was purified as shown in Figure 196, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1822

A DNA sequence (GBSx1929) was identified in *S.agalactiae* <SEQ ID 5661> which encodes the amino acid sequence <SEQ ID 5662>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 2920:

Identities = 64/91 (70%), Positives = 78/91 (85%)

Query: 1 MSDGKADFKLFDGPTVNAIKNQGLTNLKTIPLTMRDQPYIYFIFGQDQKDLQKYVNNRL 60

+S+GKADFK+FD PTVNAIKNQGL NLKTI LT +QP+IYFIF QDQ+ LQ +VN R+

Sbjct: 187 LSEKADFKIFDAPTVNAIKNQGLDNLKTIELTSTEQPFYIFIFSQDQEKLSFVNKRI 246

Query: 61 KQLRKDGTLSKIAKEYLGGDYVPNEKDLVTP 91

K+L DGTLSK+AKE+LGGDYVP++K+L P

Sbjct: 247 KELTADGTLSKLAKHEHLGGDYVPSDKELKLP 277

-2054-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1823

A DNA sequence (GBSx1930) was identified in *S.agalactiae* <SEQ ID 5663> which encodes the amino acid sequence <SEQ ID 5664>. This protein is predicted to be 28 kDa outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.44    Transmembrane    25 - 41 ( 25 - 42)

----- Final Results -----
          bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59825 GB:AJ012388 hypothetical protein [Lactococcus lactis]
Identities = 110/283 (38%), Positives = 175/283 (60%), Gaps = 13/283 (4%)

Query: 22  KLKHIVLGLALTTLGV---TFNQEVSSASSTSSKVVKVGVMTFSDTEKARWDKIEKLV 77
           K ++I++ +A+ L+ +   + ++Q   +S  K VKVG+M+   ++ W  +
Sbjct: 4   KNRNIIIAVAVLILVALVAFFSLNHQGGVKASAGEKTVKVGIMSGDKQDQEVWKS VANTA 63

Query: 78  GDK--AKIKFTEFTDYTOPNQATANKDVIDINAFQHYNFLENWKNENKKNLIPLKTYLAP 135
           +K  K+KF  F+DY QPN+A  + D+DINAFQ YN+++ WNK +K +++ +  TY+ P
Sbjct: 64  KEKYDLKLKFPVYFSDYNQPNALLSGDIDINAFQSYNYVKTWNKAHKSDIVAVGNTYITP 123

Query: 136 IRIYSEKVKSLKKLKGATTAIPNDATNGSRALYVLQSAGLIKLNVS-GKKVATVANITS 194
           + IYS+++ L  LK+G+T+AIPNDA+N SRAL+VLQSAGL+KL  S  K+  + +IT
Sbjct: 124 MHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQSAGLLKLTSDSSKLVGLPDITE 183

Query: 195 NKKDINIQELDASQTPRALKDVDAAIINNNTYIEQANLKPSDAIFVEKSDKNSKQWINIIA 254
           N  +  +E+DASQTPRAL V +++N Y  A+L S+++F+E  +K S Q+IN IA
Sbjct: 184 NPHQLKFKEVDASQTPRALDSVALSVVNYNYATAASLPKSESVFMEPLNKTSAQYINFIA 243

Query: 255 GRKNWKKQKNAKAIQAILDAYHTDEVKKVIKDTSDAD---IPQW 294
           K+KN K  + +  AY +   +K IK+  D   +P W
Sbjct: 244 ---TTSKEKNKVKYKEVAKAYASKATEKAIKEQYPDGGELPAW 283
```

There is also homology to SEQ ID 2132.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8907> and protein <SEQ ID 8908> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1    Crend: 4
McG: Discrim Score:      7.47
GvH: Signal Score (-7.5): -4.79
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
ALOM program  count: 1 value: -1.44 threshold: 0.0
    INTEGRAL    Likelihood = -1.44    Transmembrane    5 - 21 ( 5 - 22)
    PERIPHERAL  Likelihood = 5.20      147
modified ALOM score: 0.79

*** Reasoning Step: 3

----- Final Results -----
```

```

bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

5 The protein has homology with the following sequences in the databases:

40.6/63.1% over 279aa

Lactococcus lactis

GP|6165402| hypothetical protein Insert characterized

```

10      ORF00442(364 - 1182 of 1482)
      GP|6165402|emb|CAB59825.1||AJ012388(4 - 283 of 287) hypothetical protein {Lactococcus
      lactis}
      %Match = 21.0
      %Identity = 40.6 %Similarity = 63.0
15      Matches = 112 Mismatches = 96 Conservative Sub.s = 62

```

162 192 222 252 282 312 342 372
WDTFKNS*RI~~PWR~~*LRTK*ERSRYS*GEVVIKTKEMSILSFLLYSLKL*QETVYNNLILITSYGIISLSQKLREFIMKLK
| :
MNPKNR

25

| | | | | | | | | | | | | | | |
|--|---|----|------|----|-----|----|---------|----|------|----|--------------|----|-----|----|
| | 402 | | 450 | | 480 | | 510 | | 540 | | 564 | | 594 | |
| | HIVLGLALTTLLG--VTFS--NQEVSASSTSSKVVVKGVMTFSDTEKARWDKIEKLVGDK--AKIKFTFTDYTQPNQAT | | | | | | | | | | | | | |
| | : :: : : : | | : : | | : : | | :: : : | | : : | | : : : | | | |
| | NIIIAVAVLILVALVAFFSLNHQGGSAGETVKVKGIMSGDKQDQEVWKS VANTAKEKYDLKLKFVYFSDYNQNPEAL | | | | | | | | | | | | | |
| | | 20 | | 30 | | 40 | | 50 | | 60 | | 70 | | 80 |

ANKDVDINAFQHYNFLENWNNKKNLIPLEKTYLAPIRIYSEKVSKLKLLKKGATTAIPNDATNGSRALYVLQSAGLIK
:
LSGDIDINAFQSYNYVKTWNKAHKSDIVAVGNTYITPMHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQSAGLLK

35

| | | | | | | | | | |
|--|--|-----|-----|-----|-----|------|------|------|--|
| | 861 | 891 | 921 | 951 | 981 | 1011 | 1041 | 1071 | |
| | LNVS-GKKVATVANITSNKKDINIQELDASQTFRALKVDAAIINNNTYIEQANLKPSDAIFVEKSDKNKSKOWINITIAGRK | | | | | | | | |
| | : : : : : :: : ::: : : | | | | | | | | |
| | LTTSDDSKLVGLPDITENPHQLKFKEVDASQTFRALDSVALSVNNYNATAAALPKSESVFMPELNKTSQAQYINFIA--- | | | | | | | | |
| | 180 | 190 | 200 | 210 | 220 | 230 | 240 | | |

```

40      1101      1131      1161      1182      1212      1242      1272      1302
NWKKQKNAKAIQAILDAYHTDEVKKVIKDTSD---IPQW*RELT*V*QGILIGYNLSAI*P*RAWDEYNVPGSWIVFE
  | : | | : : | | : : | | : | : | : |
TTSKEKNKNKVYKEVAKAYASKATEKAIKEQYPDGGELPAWDLKL
      260      270      280

```

45 SEQ ID 8908 (GBS35) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 2; MW 31.6kDa).

The GBS35-His fusion product was purified (Figure 96A; see also Figure 192, lane 6) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 96B), FACS (Figure 96C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 1824

A DNA sequence (GBSx1931) was identified in *S.agalactiae* <SEQ ID 5665> which encodes the amino acid sequence <SEQ ID 5666>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
55 >>> Seems to have no N-terminal signal sequence

```

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.3126(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

-2056-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF11560 GB:AE002038 ArgE/DapE/AcyI family protein [Deinococcus radiodurans]
Identities = 129/419 (30%), Positives = 210/419 (49%), Gaps = 14/419 (3%)

Query: 26 LRLDIAIKSIFAQKVGLNDLSSYLGEVFIKAGAEVIIDDSYSAFFIVANFKSSKVDARI 85
LR L+A+ S+ AQ L + + + + G V AP ++A +

10 Sbjct: 16 LRALVALPSVSAQGRMLPETADAVAGLLRAEGFGVQQFPGTVPVLLAEAGEGPFT---L 72

Query: 86 IFYNHYDTVPADEVEQWTEDPFTLSLRVGYKMYGRGVDDDKGHITARLSAVKKYLSRHKG 145
+ YNHYD P D +E W PF L+ R G++YGRG DDKG + +RL+AV+ + G

Sbjct: 73 LIYNHYDVQPEDPLELWDTPPFELTERGGRLYGRGASDDKGEIASRLAAVRA-VREQLGH 131

15 Query: 146 LPLDITFIVEGAEEASASVGLDYLEKYQEQLQADLIVWEDGPKNPKGQLEIAGGNKGIV 205
LP+ I +++EG EE S L+ ++ ++ +LQ AD WE G +P+G+ ++ G KG++

Sbjct: 132 LPVKIKWLEGEVEVGSPTLERFVAEHAAELQ-ADGCWWEFGGISPEGRPILSLGLKGVM 190

20 Query: 206 TFDLSVSSADVDIHSSFGGVDSSTWYLIQALNLRDNKGHILVEGIYDKVIPPTKRELE 265
+L AD D+HSS G V+D+ + L +A+ +LRD +G++ + G YD V + + +

Sbjct: 191 CLELRRCRVADSDLHSSLGAVIDNPLYCLARAVASLRDEQGNVTIPGFYDDVRAASGADRQ 250

Query: 266 LVEKYSYRSKALEGAYQLVLP SLADSHKTFRLKLYFEPSIAIEGITSQYQGEVKTILP 325
+ + +A+ + + P + + + P + + G GYQGEK KT+LP

25 Sbjct: 251 AIAQIP-GDGQAVRDTFGVRRP--LATGPAYNERTNLHPVVNVNGWGGGYQGEKSKTVLP 307

Query: 326 AYAKCKAEVRLVPLTPKGVLDLSIQNHLKENGFKDIELT-YTLGEMSYRSDMSAPSILKV 384
K + RLVP P VL ++ HL G DIE+ + R+D P +

30 Sbjct: 308 GAGFVKLDFRLVPDQDPARVLSLREHLTAQGLSDIEVVELEAHQKPARADAGHPFVQAC 367

Query: 385 VDLAEQFYPEGISLLPTSPGTGPMY-----LVHQALRAPIAAIGIGHANSRDHGVDENV 438
V A + + + P+S +GPM+ L . P A+GIG+ R H +EN+

Sbjct: 368 VAAARAAHGQDPPIVHPSSGASGPMFPFTGGAGGGGLGIPCVAVGIGNHAGRVHAPNENI 426

35 There is also homology to SEQ ID 2588.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1825

40 A DNA sequence (GBSx1932) was identified in *S.agalactiae* <SEQ ID 5667> which encodes the amino acid sequence <SEQ ID 5668>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.5366(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB59828 GB:AJ012388 hypothetical protein [Lactococcus lactis]
Identities = 187/338 (55%), Positives = 256/338 (75%), Gaps = 12/338 (3%)

55 Query: 6 IIKLDNIDVTFHQKKREINAVKDVITHINQGDYIGVYSGAGKSTLVRVINLLQEPSAG 65
II+L+N+ V FHQK R + AVK+ T+HI +GDIYG++GYSGAGKSTLVR INLLQ+P+ G

Sbjct: 4 IIELNNLSVQFHQKGRVLTAVKNATLHIEKGDIYGVYSGAGKSTLVRTINLLQKPTG 63

Query: 66 KITIDDQVIYD--NKVILTSTQLREQRREIGMIFQHFNLSQLTAEQNVAFALKHSG--- 120
+I I+ + I+D N V T +LRE R++IGMIFQHFNLS+ T NVAFAL+HS

-2057-

Sbjct: 64 QIVINGEKIFDSENPVKFTGAKLREFRQKIGMIFQHFNLLSEKTVFNNVAFALQHSQIED 123

Query: 121 -----LSKEAKAAKVAKILLEVLSDRAQNYPQLSGGQKQRVAIARALANDPKILIS 173
L+K+ K KV +LL+LV L+D + YP+QLSGGQKQRVAIARALANDP+ILIS

5 Sbjct: 124 KNGKRYLTKKEKNDKVTLELLKLVDLADLSDKYPALSGGQKQRVAIARALANDPEILIS 183

Query: 174 DESTSALDPKTTKQILALLQDINKKLGITVLIITHEMQIVKDIANRVAVMONGKLIIEGS 233
DE TSALDPKTT QIL LL+ L++KLG+T+VLIITHEMQ+VK+IAN+VAVMONG++IE+ S

10 Sbjct: 184 DEGTSALDPKTTNQILDLLKSLHEKLGITVVLITHEMQVVKIANKVAVMONGEIEEQNS 243

Query: 234 VLDIFSHPRESLTQDFIKIATGIDEAMLKIEQQEVVKNLPVGSKLVLQKYAGHSTDEPLL 293
++DIF+ P+E+LT+ FI+ + ++ + + + E++ L +L+ L Y+G ++P++

Sbjct: 244 LIDIFAQPKALTKQFIETTSSVNRFIASLSKTELLAQLADDEELIHLDYSGSELEDPVV 303

15 Query: 294 NQIYKEFEVTANILYGNIEILDGIPVGEMVVILSGDEE 331
+ I K+F+VT NI YGN+E+L G P G +V+ L G E

Sbjct: 304 SDITKKFDVTTNIFYGNVELLQGQPFGLVLTGKSSE 341

There is also homology to SEQ ID 76.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1826

- A DNA sequence (GBSx1933) was identified in *S.agalactiae* <SEQ ID 5669> which encodes the amino acid sequence <SEQ ID 5670>. This protein is predicted to be ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -12.79 Transmembrane 203 - 219 (197 - 225)
INTEGRAL Likelihood = -8.86 Transmembrane 73 - 89 (69 - 102)
INTEGRAL Likelihood = -7.38 Transmembrane 38 - 54 (35 - 56)
INTEGRAL Likelihood = -1.12 Transmembrane 103 - 119 (103 - 119)

----- Final Results -----

35 bacterial membrane --- Certainty=0.6116(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10083> which encodes amino acid sequence <SEQ ID 10084> was also identified.

- 40 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB59829 GB:AJ012388 hypothetical protein. [Lactococcus lactis]
Identities = 137/231 (59%), Positives = 171/231 (73%), Gaps = 1/231 (0%)

45 Query: 1 MIEWIQTHLPNVYQMGWEGAYGWQTAIVQTLTYMTFWSFLIGGLMGLLGLFLVLTSPRGV 60
M EW PNV +GW G GW TAIVQTLTYMTF S LIGGL+GL+ G+ +V+T+ G+
Sbjct: 1 MAEWFHTFPNVVYLGWTGETGWWTAIVQTLTYMTFISALIGGLGLIFGIGVVVTAEDGI 60

Query: 61 IANKLVFGVLDKVVSVFRALPFIILLALIAPVTRVIVGTTLGSPAALVPLSLAVFPFFAR 120
N+ +F +LDK+VS+ RA PFIILLA IAP+T+++VGT +G AALVPL+L V PF+AR

50 Sbjct: 61 TPNRPLFWILDKIVSIGRAFFFIILLAAIAPLTIKILVGTQIGVTAALVPLALGVAPFYAR 120

Query: 121 QVQVLAELDGGVIEAAQASGGTLWDII-VVYLREGLPDLIRVSTVTLISLVGETAMAGA 179
QVQ L +D G +EAAQ G DI+ VYLRE L LIRVSTVTLISL+G TAMAGA

Sbjct: 121 QVQASLESVDHGKVEAAQTVGADFLDIVFTVYLREELASLIRVSTVTLISLIGLTAMAGA 180

55 Query: 180 IGAGGLGSVAITKGYNYSRDDITLVATILILLILFFIQFLGDFLTRRLSHK 230
IGAGGLG+ AI+ GYN +D+T ATILIL+ + +Q +GDFL RR+SH+

Sbjct: 181 IGAGGLGNTAISYGYNRFANDVTWFATILILIFVLLVQLVGDFLARRVSHR 231

-2058-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5671> which encodes the amino acid sequence <SEQ ID 5672>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood =-11.15    Transmembrane  194 - 210 ( 187 - 215)
INTEGRAL    Likelihood =-10.67    Transmembrane  28 - 44 ( 20 - 52)
INTEGRAL    Likelihood = -8.12    Transmembrane  70 - 86 ( 62 - 91)

----- Final Results -----
bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAB59829 GB:AJ012388 hypothetical protein [Lactococcus lactis]
Identities = 123/213 (57%), Positives = 153/213 (71%), Gaps = 1/213 (0%)

Query: 9   GDAGWGLAIWNTLYMTIVPFIVGGAIGLLGLLLVLTGPDGVIENTICWVIDKVTISIFR 68
Sbjct: 19  GETGWWTAIVQTYLMTFISALIGLLGLLIFGIGVVVTAEDGITPNRPLFWILDKIVSIGR 78

Query: 69  AIPFVILIAILASFTYLLRLTLTGATAALVPLTFATFPFYARQVQVVFSELDKGVIEAAQ 128
Sbjct: 79  A PF+IL+A +A T +L+ T +G TAALVPL PFYARQVQ +D G +EAAQ

Query: 129 ASGATFWDIV-KVYLSEGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGNVAISYGYNRF 187
Sbjct: 139 TVGADFLDIVFTVYLREELASLIRVSTVTLISLIGLTAMAGAIGAGGLGNTAISYGYNRF 198

Query: 188 NNDVTWVATIIILLIIFAIQFIGDSLTRRFSSH 220
Sbjct: 199 NDVTW ATI+IL+ + +Q +GD L RR SH+

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 146/212 (68%), Positives = 172/212 (80%)

Query: 19  GAYGWQTAIVQTYLMTFWSFLIGGLMGLLGLFLVLTSPRGVIANKLVFGLDKVVSVFR 78
Sbjct: 9   GDAGWGLAIWNTLYMTIVPFIVGGAIGLLGLLLVLTGPDGVIENTICWVIDKVTISIFR 68

Query: 79  ALPFIILLALIAIPVTRVIVGTTLGSPAALVPLSLAVFPFFARQVQVVLAEELDGGVIEAAQ 138
Sbjct: 69  A+PF+IL+A++A T +++ TTLG+ AALVPL+ A FPF+ARQVQVV +ELD GVIEAAQ

Query: 139 ASGGTLWDIIVVYLREGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGSAITKGYNSR 198
Sbjct: 129 ASGATFWDIVKVYLSEGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGNVAISYGYNRFN 188

Query: 199 DDITLVATILILLIIFIQFLGDFLTRRLSHK 230
Sbjct: 189 NDVTWVATIIILLIIFAIQFIGDSLTRRFSSH 220

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1827

A DNA sequence (GBSx1934) was identified in *S.agalactiae* <SEQ ID 5673> which encodes the amino acid sequence <SEQ ID 5674>. This protein is predicted to be alcohol dehydrogenase, zinc-containing (Zn-dependent). Analysis of this protein sequence reveals the following:

-2059-

Possible site: 21

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.92 Transmembrane 71 - 87 (69 - 87)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.2168(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 9419> which encodes amino acid sequence <SEQ ID 9420> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF41759 GB:AE002488 alcohol dehydrogenase, zinc-containing

[Neisseria meningitidis MC58]

15 Identities = 135/246 (54%), Positives = 186/246 (74%), Gaps = 1/246 (0%)

Query: 3 SHCEDGGWILGHLIEGTQAEYVHIPHADGSLYHAPEGVCDDALVMSDILPTS YEIGVLP 62

SHC +GGWILG++I+GTQAEYV P+AD SL P+ V ++ ++LSD LPT++EIGV

20 Sbjct: 102 SHCRNGGWILGYMIDGTQAEYVTPYADNSLVPLPDNVNEEIALLLSDALPTAHEIGVQY 161

Query: 63 SHIKPGDTCIVGAGPIGLSALLTAQFYSPAKIIMVDLSQKRLEASKKFGATHTILSTST 122

+KPGDTV I GAGP+G+SALLTAQ YSPA II+ D+ + RL+ +K+ GATHTI + ++

Sbjct: 162 GDVKGPDTCVFIAGAGPVGMSALLTAQLYSPAIIIVCDMDENRLKLAKELGATHTI-NPAS 220

25 Query: 123 QEVKEEIDKITKGRGVDVVLECVGYPATFDICQNVVSIGGHIANVGVHGKPVFENLQDLW 182

EV +++ I GVD +E VG PAT+++CQ++V GGHIA VGVHG+ V+F L+ LW

Sbjct: 221 GEVSKQVFAIVGEDGVDCAEAVGIPATWNMCQDIVKPGGHIAVVGHVHGQSVDFKLEKLW 280

30 Query: 183 IKNITLNTGLVNANTTEMLLEVELETGKIDATQLVTHHFKLSEIEEAYKVFKAEEENNTLK 242

IK + + TGLVNANTTEML++ + + +D T+++THHFK SE+E+AY VFK A EN +K

Sbjct: 281 IKKLAITTLGLVNANTTEMLMKAISSSSVDYTKMLTHHFKFSELEKAYDVFKHAAENQVMK 340

Query: 243 VIIEND 248

V++E D

35 Sbjct: 341 VVLEAD 346

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 785> which encodes the amino acid sequence <SEQ ID 786>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -5.41 Transmembrane 184 - 200 (183 - 203)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.3166(Affirmative) < succ>

45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 199/250 (79%), Positives = 226/250 (89%)

50 Query: 1 MPSCEDGGWILGHLIEGTQAEYVHIPHADGSLYHAPEGVCDDALVMSDILPTS YEIGV 60

+ SHC+DGGWILGHLI GTQAEYVHIPHADGSLYHAP+ + D+ALVMSDILPTS YEIGV

Sbjct: 114 LSSHCQDGGWILGHLINGTQAEYVHIPHADGSLYHAPDTIDDEALVMSDILPTS YEIGV 173

55 Query: 61 LP SHIKPGDTCIVGAGPIGLSALLTAQFYSPAKIIMVDLSQKRLEASKKFGATHTILST 120

LP SH+KPGD VCIVGAGP+GL+ALLT QF+SPA IIMVDLSQ RLEA+K FGATHTI S

Sbjct: 174 LP SHVKGPDNVCIVGAGPVGLAALLTVQFFSPANIIMVDLSQNRLEAAKTFGATHTICSG 233

60 Query: 121 STQEVKEEIDKITKGRGVDVVLECVGYPATFDICQNVVSIGGHIANVGVHGKPVFENLQD 180

S++EVK ID IT GRGVD+ +ECVGYPATFDICQ ++S+GGHIANVGVHGKPV+FNL +

Sbjct: 234 SSEEVKAIIDDIITNGRGVDISMECVGYPATFDICQKIISVGGHIANVGVHGKPVDFNLDE 293

-2060-

Query: 181 LWIKNITLNTGLVNANTTEMLLEVLETGKIDATQLVTHHFKLSEIEAYKVFKAAEENNT 240
 LWIKNITLNTGLVNANTTEMLL VL+TGKIDAT+L+THHFKLSE+E+AY+ FK A NN
 Sbjct: 294 LWIKNITLNTGLVNANTTEMLLNVLKTGKIDATRLITHHFKLSEVEKAYETFKHAGANNA 353

5 Query: 241 LKVIIENDIT 250
 LKVII+NDI+
 Sbjct: 354 LKVIIDNDIS 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1828

A DNA sequence (GBSx1935) was identified in *S.agalactiae* <SEQ ID 5675> which encodes the amino acid sequence <SEQ ID 5676>. This protein is predicted to be a dehydrogenase fragment. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.46 Transmembrane 47 - 63 (33 - 66)

----- Final Results -----
 bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 786:

Identities = 23/38 (60%), Positives = 28/38 (73%)

Query: 7 WRNSNMRAATYLSANELSLTDKAKPQVIKPTDAVVLV 44
 ++ NM+AAATYLS L L DK KP +IKPTDA+V LV
 Sbjct: 10 YKCLNMKAATYLSGNLQLIDKPKPVIKPTDAIVQLV 47

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1829

A DNA sequence (GBSx1936) was identified in *S.agalactiae* <SEQ ID 5677> which encodes the amino acid sequence <SEQ ID 5678>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2061-

Example 1830

A DNA sequence (GBSx1937) was identified in *S.galactiae* <SEQ ID 5679> which encodes the amino acid sequence <SEQ ID 5680>. This protein is predicted to be branched chain amino acid transport system II carrier protein (brnQ). Analysis of this protein sequence reveals the following:

```

5   Possible site: 44
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.66    Transmembrane 158 - 174 ( 154 - 177)
      INTEGRAL    Likelihood = -6.64    Transmembrane 233 - 249 ( 231 - 252)
      INTEGRAL    Likelihood = -5.20    Transmembrane 37 - 53 ( 30 - 57)
10   INTEGRAL    Likelihood = -3.98    Transmembrane 90 - 106 ( 87 - 108)
      INTEGRAL    Likelihood = -0.80    Transmembrane 130 - 146 ( 130 - 146)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
15   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9417> which encodes amino acid sequence <SEQ ID 9418> was also identified.

20 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAC00400 GB:AF008220 branch-chain amino acid transporter
    [Bacillus subtilis]
    Identities = 89/250 (35%), Positives = 139/250 (55%), Gaps = 18/250 (7%)

25   Query: 1   MDALASIAFAIIVIQASKQYGAIKKKEITSMALKSGAIATFLLAFIYIFVGRIGATSQSL 60
      MDALASI F ++V+ A K G   K + + +K+G IA   L FIY+ + +GATS +
    Sbjct: 199 MDALASIVGVVVVNAVSKSGVTQSKALAAACIKAGVIAALGLTFIYVSLAYLGATSTNA 258

    Query: 61   FKFANGSFLHNTPI-DGGHVLSQSANFYLGIVGQAILGTAIFLACLTTATGLITACA EY 119
      P+ +G +LS S+++ G +G +LG AI +ACLTT+ GL+T+C +Y
30   Sbjct: 259 IG-----PVGEGAKILSASSHYLFGSLGNIVLGAATVACLTTSIGLVTSCGQY 307

    Query: 120  FHKLLPKISHITWATIFTLIAITFYFGGLSEIIRWSLPVLYLLYPLTIVLIFLVFFDQKF 179
      F KL+P +S+   TI TL ++   GL++II +S+P+L +YPL IV+I L F D+ F
35   Sbjct: 308 FSKLIPALSYKIVVTIVTLFSLIIANFGLAQIIAFSVPILSAIYPLAIVIIIVLSFIDKIF 367

    Query: 180  ESSRIVYQTSIAATAVAALYDALSKLGEMTGLFTIPSALTTFFTKVVPLGEYSMGWISFA 239
      + R VY   + T + ++ D +   G G   +L F   +PL   +GW+
40   Sbjct: 368 KERREVIYIACLIGHTGLFSILDGIKAAGFSLG-----SLDVFLNANLPLYSLGIGWVLP 421

    Query: 240  ICGVLVGLIL 249
      I G ++G +L
    Sbjct: 422 IVGAVIGYVL 431

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2233> which encodes the amino acid sequence <SEQ ID 2234>. Analysis of this protein sequence reveals the following:

```

    possible site: 21
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -10.83   Transmembrane 235 - 251 ( 228 - 258)
50   INTEGRAL    Likelihood = -8.49    Transmembrane 434 - 450 ( 429 - 454)
      INTEGRAL    Likelihood = -8.12    Transmembrane 359 - 375 ( 356 - 377)
      INTEGRAL    Likelihood = -7.86    Transmembrane 150 - 166 ( 144 - 171)
      INTEGRAL    Likelihood = -6.00    Transmembrane 298 - 314 ( 288 - 316)
      INTEGRAL    Likelihood = -5.95    Transmembrane 42 - 58 ( 38 - 63)
55   INTEGRAL    Likelihood = -3.35    Transmembrane 336 - 352 ( 335 - 354)
      INTEGRAL    Likelihood = -2.81    Transmembrane 199 - 215 ( 198 - 218)
      INTEGRAL    Likelihood = -2.18    Transmembrane 120 - 136 ( 120 - 138)
      INTEGRAL    Likelihood = -1.81    Transmembrane 390 - 406 ( 390 - 407)
      INTEGRAL    Likelihood = -1.01    Transmembrane 81 - 97 ( 81 - 97)
60

```

-2062-

----- Final Results -----

bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 161/253 (63%), Positives = 197/253 (77%)

Query: 1 MDALASIAFAIIVIQASKQYGAITKKEITSMALKSGAIATFLLAFIYIFVGRIGATSQSL 60
 MDALAS+ FAI+VI+A+KQ+GA T KE+T + L SGAIA LLA +YIFVGRIGATSQSL
 Sbjct: 202 MDALASLVFAILVIEATKQFGAKTDKEMTKITLISGAIAILLALLLVYIFVGRIGATSQSL 261

Query: 61 FKFANGSFLHNTPIDGGHVSQSANFYLGIVGQAILGTAIFLACLTTATGLITACAEYF 120
 F F +GSF LH P++GG +LS ++ FYLG +GQA L IFLACLTT+TGLIT+ AEYF
 Sbjct: 262 PPFIDGSFTLHGNPVNGGQILSHASRFYLGIGQAFLAVVIFLACLTTSTGLITSSAEYF 321

Query: 121 HKLLPKISHITWATIFTLIAITFYFGGLSEIIRWSLPVLYLLYPLTIVLIFLVFFDQKFE 180
 HKL+P +SHI WATIFTL++ FYFGGLS II WS PVL+LLYPLT+ LIFLV + F
 Sbjct: 322 HKLVPALSHIAWATIFTLLSAFFYFGGLSVIINWSAPVLFLLYPLTVDLIFLVLAQKCFN 381

Query: 181 SSRIVYQTSIAATAVAALYDALSKLGEMTGLFTIPSALTTFFTKVPLGEYSMGWISFAI 240
 + IVY+T+I T + A++DAL L +MTGLE +P A+ TFF K VPLG++SMGWI FA
 Sbjct: 382 NDPIVYRTTIGLTFIPAIFDALLTSLQMTGLFHLPEAVVTFQKTVPLGQFSMGWIIFAA 441

Query: 241 CGVLVGLILKKVK 253
 G L+GLIL K K
 Sbjct: 442 IGFLIGLILSKTK 454

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1831

A DNA sequence (GBSx1938) was identified in *S.agalactiae* <SEQ ID 5681> which encodes the amino acid sequence <SEQ ID 5682>. This protein is predicted to be 30S ribosomal protein S12 (rpsL). Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9429> which encodes amino acid sequence <SEQ ID 9430> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA78825 GB:Z15120 ribosomal protein S12 [Streptococcus pneumoniae]
 Identities = 64/71 (90%), Positives = 68/71 (95%)

Query: 1 MPTINQLVRKPRKSKVEKSDSPALNIGYNHRKVHTKLSAPQKRGVATRVGTMTPKKPNS 60
 MPTINQLVRKPRKSKVEKS SPALN+GYNSH+KV T +S+PQKRGVATRVGTMTPKKPNS
 Sbjct: 1 MPTINQLVRKPRKSKVEKSKSPALNVGYNHKKVQTNVSSPQKRGVATRVGTMTPKKPNS 60

Query: 61 ALRKVFARVRLS 71
 ALRKVFARVRLS

Sbjct: 61 ALRKVFARVRLS 71

-2063-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5683> which encodes the amino acid sequence <SEQ ID 5684>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3879(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 44/48 (91%), Positives = 47/48 (97%)

Query: 24 LNIGYNShrkvhtklsapQKRGVATRVGTMTPKKPNSALRKfARVRLS 71

LNIGYNsh+kv TK++APQKRGVATRVGTMTPKKPNSALRKfARVRLS

Sbjct: 1 LNIGYNshkkvQTKMAAPQKRGVATRVGTMTPKKPNSALRKfARVRLS 48

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1832

A DNA sequence (GBSx1939) was identified in *S.agalactiae* <SEQ ID 5685> which encodes the amino acid sequence <SEQ ID 5686>. This protein is predicted to be purR. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -0.37 Transmembrane 142 - 158 (142 - 159)

----- Final Results -----

bacterial membrane --- Certainty=0.1150(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10902 GB:AJ222642 purR [Lactococcus lactis]

Identities = 143/269 (53%), Positives = 195/269 (72%), Gaps = 1/269 (0%)

Query: 3 LRRSERMVVISNYLINPYTLTSLNTFASKYGAAKSSISEDIAIIKKAFAEQAIQIGDIKTIV 62

++R+ER+V +N+LIN+P + +LN + Y AKSSISED+ IK+ FE +G ++T

Sbjct: 1 MKRNERLVDFTNFLINHPNQMLNLNELSKHYEVAKSSISEDLVFIKRVFENQGVGLVETF 60

Query: 63 TGASGGVIFTPTIAEAEAKEIVEELRQRLSENDRIPLGGYIYLSDLLSTPKMLQSIGRII 122

G+ GGV FTP I + + E+ +E+ + L E +RILPGGYIYLSL+L TP L+ IG+II

Sbjct: 61 PGSLLGGVRFTPYITDERSLEMSQEIABLLREENRILPGGYIYLSLILGTPSNLRKIGQII 120

Query: 123 ANAYRGQKIDAVMTVATKGVPLANAVANVLDVPFVIVRRDLKITEGSTVSVNYASGSSGR 182

A+ Y +++D VMT+ATKG+P+A +VA +LDVPFVIVRRD K+TEG+T++VNY SGSS R

Sbjct: 121 AHEYHEKQVDVMTIATKGIPIAQSVAEILDVPFVIVRRDPKVTEGATLNVNMSGSSSR 180

Query: 183 IEKMFLSKRSLKPNRSRLIVDDFLKGGGTVSGMISLLSEFDSTLVGVAVFAENA-QEQRE 241

+E M LSKRSL VLIIVDDF+KG GT++GM SL+ EFD L GVAVF E + +R

Sbjct: 181 VENMTLSKRSLSIGQNVLIIVDDFMKGAGTINGMRSLVHEFDCLLAGVAVFLEGPFKGERL 240

Query: 242 KMAYKSLLRVSEIDVKNNRVSVEAGNIFD 270

YKS+L+V ID+ N + V+ GNIF+

Sbjct: 241 IDDYKSILKVDRIDIANRSIDVQLGNIFN 269

-2064-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5687> which encodes the amino acid sequence <SEQ ID 5688>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
5   INTEGRAL    Likelihood = -1.97    Transmembrane  142 - 158 ( 142 - 160)

----- Final Results -----
          bacterial membrane --- Certainty=0.1786(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAA10902 GB:AJ222642 purR [Lactococcus lactis]
Identities = 142/269 (52%), Positives = 196/269 (72%), Gaps = 1/269 (0%)
15
Query: 3   LRRSERMVVISNYLINNPYKLTSLNTFATKYEAAKSSISEDIAI IKKAFEEANIGDIDTL 62
          ++R+ER+V  +N+LIN+P ++ +LN  +  YE AKSSISED+  IK+ FE  +G ++T
Sbjct: 1   MKRNERLVDFINFLINHPNQMLNLSKHYEVAKSSISEDLVFIKRVFENQGVGLVETF 60

20
Query: 63   TGASGGVIFTPSISETEARTIVEDLCQRLSESDRILPGGYIYLSDLLSTPKILQNIIRII 122
          G+ GGV FTP I++  +  + +++ + L E +RILPGGYIYLSL+L TP  L+ IG+II
Sbjct: 61   PGS LGGVRFPTPYITDERSLEMSQBIABLLREENRILPGGYIYLSLILGTPSNLRKIGQII 120

25
Query: 123  ANAFKGEKIDAVMTVATKGVPLANAVANILSVPFVIVRRDLKITEGSTVSVNYASASSDR 182
          A+ +  +++D VMT+ATKG+P+A +VA IL VPFVIVRRD K+TEG+T++VNY S SS R
Sbjct: 121  AHEYHEKQVDVMTIATKGIPIAQSVAEILDVFPFVIVRRDPKVTEGATLNVNMSGSSSR 180

Query: 183  IEKMFSLSKRSLKPNRSRVLIVDDFLKGGGTITGMISLLTEFDSTLVGVAVFAENA-QSERE 241
          +E M LSKRSL  VLI VDDF+KG GTI GM SL+ EFD  L GVAVF E  + ER
30
Sbjct: 181  VENMTLSKRSLSIGQNVLIVDDFMKGAGTINGMRSLVHEFDCLLAGVAVFLEGPFKGERL 240

Query: 242  QMTFKSLLKVSEIDVKNNNVVVEVGNIFD 270
          +KS+LKV  ID+ N ++ V++GNIF+
35
Sbjct: 241  IDDYKSILKVDRIDIANRSIDVQLGNIFN 269

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 234/270 (86%), Positives = 255/270 (93%)
40
Query: 1   MKLRRSERMVVISNYLINNPYTLTSLNTFASKYGAAKSSISEDIAI IKKAFEAQIGDIK 60
          MKLRRSERMVVISNYLINNPY L TSLNTFA+KY AAKSSISEDIAI IKKAFE+A IGDI
Sbjct: 1   MKLRRSERMVVISNYLINNPYKLTSLNTFATKYEAAKSSISEDIAI IKKAFEEANIGDID 60

Query: 61   TVTGASGGVIFTPPTIAEAEAKEIVEELRQRLSENDRI LPGGYIYLSDLLSTPKMLQSIGR 120
          T+TGASGGVIFTP+I+E EA+ IVE+L QRLSE+DRILPGGYIYLSDLLSTPK+LQ+IGR
45
Sbjct: 61   TLTGASGGVIFTPSISETEARTIVEDLCQRLSESDRILPGGYIYLSDLLSTPKILQNIIR 120

Query: 121  IIANAYRGQKIDAVMTVATKGVPLANAVANVLDVFPFVIVRRDLKITEGSTVSVNYASGSS 180
          IIANA++G+KIDAVMTVATKGVPLANAVAN+L VPFVIVRRDLKITEGSTVSVNYAS SS
50
Sbjct: 121  IIANAFKGEKIDAVMTVATKGVPLANAVANILSVPFVIVRRDLKITEGSTVSVNYASASS 180

Query: 181  GRIEKMFLSKRSLKPNRSRVLIVDDFLKGGGTVSGMISLLSEFDSTLVGVAVFAENAQEQR 240
          RIEKMFLSKRSLKPNRSRVLIVDDFLKGGGT++GMISLL+EF DSTLVGVAVFAENAQ +R
Sbjct: 181  DRIEKMFLSKRSLKPNRSRVLIVDDFLKGGGTITGMISLLTEFDSTLVGVAVFAENAQSER 240

55
Query: 241  EKMAKSLLRVSEIDVKNNRVSV EAGNIFD 270
          E+M +KSL+VSEIDVKNN V VE GNIFD
Sbjct: 241  EQMTFKSLLKVSEIDVKNNNVVVEVGNIFD 270

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2065-

Example 1833

A DNA sequence (GBSx1940) was identified in *S.agalactiae* <SEQ ID 5689> which encodes the amino acid sequence <SEQ ID 5690>. This protein is predicted to be cmp-binding-factor 1. Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1753(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAC44803 GB:U21636 cmp-binding-factor 1 [Staphylococcus aureus]
    Identities = 140/310 (45%), Positives = 195/310 (62%), Gaps = 6/310 (1%)

    Query: 3   INQMKKDELFEGFYLIKKAIEVRKTRAGKDFIAFTFQDDTGEISGNMWDATYNVEEFVAG 62
              I + + + F+L+ KA T GKD++ QD +GEI W A ++
    Sbjct: 4   IENLNPGDSVDHFFLVHKATQGVTAQGKDYMTLHLQDKSGEIEAKFWTATKNDMATIKPE 63

20  Query: 63   KIVHMKGRREVYNGTPQ--VNQITLNRNIKDGEPNDRDFKEKPPINVDNVREYMEQMLFK 120
              +IVH+KG Y G Q VNQI L +D + F + P++ ++E + L
    Sbjct: 64   EIVHVKGDIINRGNKQMKVNQIRLATTEDQLKTE--QFVDGAPLSPAIEQEEISHYLLD 121

25  Query: 121  IENATWQVRVVRALYRKYNKEFFTYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYPELN 180
              IENA QR+ R L +KY + F+TYPAA ++HH F SGL+YH TM+R+A SI DIYP LN
    Sbjct: 122  IENANLQRITRHLKKYQERFYTYPAASSHHHNFASGLSYHVLTMLRIAKSICDIYPLLN 181

30  Query: 181  KSLMFAGIMLHDLAKVIELSGPDNTEYITIRGNLIGHISLIDEELTKILAEELNIDDTKEEV 240
              KSL+++GI+LHD+ KV ELSGP T YT+ GNL+GHIS+ +E+ + ELNI+ EE+
    Sbjct: 182  KSLLYSGIILHDIGKVVRELSGFPVATSYTVEGNLLGHISTASDEVVEARELNIEG--EEI 239

    Query: 241  TVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMDNIDANMMMTTALNRVNEGEMTNRIF 300
              +LRH+ILSHHG+LEYGSP P + EAEI+ IDNIDA M M A + ++G+ T++IF
35  Sbjct: 240  MLLRHMILSHHGKLEYGSPKLPYLKEAEILCYIDNIDARMNMFEEKAYKKTDKGQFTDKIF 299

    Query: 301  AMDNRSFYKP 310
              ++NR FY P
40  Sbjct: 300  GLENRRFYNP 309

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5691> which encodes the amino acid sequence <SEQ ID 5692>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 38
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1822(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 275/311 (88%), Positives = 300/311 (96%)

55  Query: 1   MKINQMKKDELFEGFYLIKKAIEVRKTRAGKDFIAFTFQDDTGEISGNMWDATYNVEEFV 60
      MKINQMKKD+LFEGFYLIK AEVRKTRAGKDFI+ TFQDDTGEISGN+WDAQ YNVEEF
    Sbjct: 1   MKINQMKKDQLFEGFYLIKSAIEVRKTRAGKDFISLTFQDDTGEISGNLWDAQPYNVEEFT 60

    Query: 61   AGKIVHMKGRREVYNGTPQVNQITLNRNIKDGEPNDRDFKEKPPINVDNVREYMEQMLFK 120
              AGK+V MKGRREVYNGTPQVNQITLRN++ GEPNDP+DFKEK P++V VR+Y+EQMLFK
60  Sbjct: 61   AGKVVFMMKGRREVYNGTPQVNQITLRNVRPGEPNDFKDFKEKAPVSVTEVRDYLEQMLFK 120

```

-2066-

Query: 121 IENATWQRVVRALYRKYNKEFFTYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYPELN 180
 IENATWQR+VRALYRKY+KEF+TYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYP+LN
 Sbjct: 121 IENATWQRIVRALYRKYDKEFYTYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYPDIN 180

5 Query: 181 KSLMFAGIMLHDLAKVIELSGPDNTEYTRGNLIGHISLIDEELTKILAEINIDDTKEEV 240
 KSL+FAGIMLHDLAKVIEL+GPDNTEYT+RGNLIGHISLI+EE+TK+++EL IDDTKEEV
 Sbjct: 181 KSLLFAGIMLHDLAKVIELTGPDNTEYTVRGNLIGHISLINEEITKVISLQIDDTKEEV 240

10 Query: 241 TVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMTTALNRVNEGEMTNRIF 300
 VLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMTTAL+RV+EGEMTNRIF
 Sbjct: 241 IVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMTTALSRVSEGEMTNRIF 300

Query: 301 AMDNRSFYKPN 311
 AMDNRSFYKPN
 15 Sbjct: 301 AMDNRSFYKPN 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1834

20 A DNA sequence (GBSx1941) was identified in *S.galactiae* <SEQ ID 5693> which encodes the amino acid sequence <SEQ ID 5694>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -14.59 Transmembrane 2 - 18 (1 - 22)
 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5695> which encodes the amino acid sequence <SEQ ID 5696>. Analysis of this protein sequence reveals the following:

Possible site: 17
 35 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.05 Transmembrane 3 - 19 (1 - 26)
 ----- Final Results -----
 40 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 309/424 (72%), Positives = 370/424 (86%), Gaps = 3/424 (0%)
 45 Query: 1 MLVILIIIVLASLTVTIIISYQKMTLTKSVEKQLEDNADNLSQDQTYQIEVAQKDQILTL 60
 +++ +L++VL L ++ K+ L + + LE NADNLSQD+TYQ++ A K Q+L L
 Sbjct: 3 LILFLVLVLGLGAYLLF--KVNGLQHQAQTLQEGNADNLSQDQTYQLDTANKQQLLEL 60

50 Query: 61 TNQLNRMQQEIIYQLLTDMRTELNQHLLTESRDRSDKRLELINSNLSQSVQKMQDSNEKRLD 120
 T +NR Q +YQ LTD+R L++ L++SRDRSDKRLE IN ++QS++ MQ+SNEKRL+
 Sbjct: 61 TQLMNRQQAGLYQQLTDIRDVLHRSLSRDRSDKRLEKINQQVNQSLKNMQESNEKRL 120

55 Query: 121 QMRQTVEEKLEKTLQTRLQTSFETVSRQLESVNQGLGEMKTVAQDVGTLNKLVSNTKTRG 180
 +MRQ VEEKLE+TL+ RL SF++VS+QLESVN+GLGEM++VAQDVGTLNKLVSNTKTRG
 Sbjct: 121 KMRQTVEEKLEETLKNRLHASFDSVSKQLESVNKGLGEMRSVAQDVGTLNKLVSNTKTRG 180

Query: 181 ILGELQLGQIIEDIMTVSQYEREFPTVSGSSERVEYAIKLPNGQGQDYIYLPIDSKFPLE 240
 ILGELQLGQIIEDIMT SQYEREF TVSGSSERVEYAIKLPNGQGQ YIYLPIDSKFPLE
 60 Sbjct: 181 ILGELQLGQIIEDIMTSSQYEREFVTVSGSSERVEYAIKLPNGQGQGYIYLPIDSKFPLE 240

-2067-

Query: 241 DYYRLEDAYELGDKVQIELYRKSLASIRKFAKDINNKYLNPPETTNGIMFLPTEGLYS 300
 DYYRLEDAYE+GDK+ IE RK+LLA+I++FAKDI+ KYLNPPETTNG+MFLPTEGLYS
 Sbjct: 241 DYYRLEDAYEVGDKLAIEASRKALLAAIKRFAKDIHKKYLNPPETTNGVGMFLPTEGLYS 300

5

Query: 301 EVVRNATFFDSLRRDENIVVAGPSTLSALLNSLSVGFKTLNIQKNANDISKILGNVKVEF 360
 EVVRNA+FFDSLRR+ENIVVAGPSTLSALLNSLSVGFKTLNIQKNA+DISKILGNVK+EF
 Sbjct: 301 EVVRNASFFDSLRRREENIVVAGPSTLSALLNSLSVGFKTLNIQKNADDISKILGNVKLEF 360

10

Query: 361 GKFGGMLSQAQKQLNTASKSIDSLLTTRTNAIRVLTVEEHQDQATTSLLNLPITEEEE 420
 KFGG+L+KAQKQ+NTA+ ++D L++TRTNAI+R LNTVE +QDQAT SLLN+P+ EEE
 Sbjct: 361 DKFGGLLAKAQKQMNTANNTLDQLISTRINAIVRALNTVETYQDQATKSLLNMPLEEN 420

15

Query: 421 INEN 424
 NEN
 Sbjct: 421 -NEN 423

SEQ ID 5694 (GBS88) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 2; MW 48kDa).

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1835

A DNA sequence (GBSx1942) was identified in *S.agalactiae* <SEQ ID 5697> which encodes the amino acid sequence <SEQ ID 5698>. Analysis of this protein sequence reveals the following:

25 Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2722(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:CAB13453 GB:Z99112 yloS [Bacillus subtilis]
 Identities = 75/217 (34%), Positives = 109/217 (49%), Gaps = 12/217 (5%)

Query: 1 MTKIALFAGG-----DLTYFEYDFDYFVGIDRGSFLFLKNGLSLDMAGVDFDSDITEDEL 54
 M I + AGG DLT + + ++G+D+G++ LL G+ A GDFDSITE E
 Sbjct: 1 MKTINIVAGGPKNLIPDLTGYTDEHTLWIGVDKGTVTLTDAGIIPVEAFGDFDSDITEQER 60

40

Query: 55 LYIKHYCSNIVSASAEKNDTDELALKTIFKEFPEAQVTVFGAFGGRIDHMSNIFLPSD 114
 I+ + AEK+ TD +LAL ++ P+ + +FG GGR DH + NI L
 Sbjct: 61 RRIEKAAPALHVYQAEKDQTDLDLALDWALEKQPDIIQIFGITGGRADHFLGNIQLLYK 119

45

Query: 115 RDLEPFMSQIRLKDEQNIVTYLPSGKNQVSRIEGMSYVSFMPSES--TLQISGAKYELN 172
 +IRL D+QN + P G+ + + E Y+SF+P SE L ++G KY LN
 Sbjct: 120 GVKTNI--KIRLIDQNHIQMFPPEYDIEKDENKRYISFIPFSEDIHELTLTGFKYPLN 177

50

Query: 173 KSNY-FKKKMYSSNEFMTSPIEVELKDGYLIIIIYSKD 208
 + + SNE + S G LI+I S D
 Sbjct: 178 NCHITLGSTLCISNELIHSRGTFSAKGILIMIRSTD 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5699> which encodes the amino acid sequence <SEQ ID 5700>. Analysis of this protein sequence reveals the following:

55 Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2068-

bacterial cytoplasm --- Certainty=0.2467(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 130/208 (62%), Positives = 166/208 (79%)

```

Query: 1  MTKIALFAGGDLTYFEYDFDYFVGIDRGSLLKNGLSLDMVAGDFDSITEDELLYIKHY 60
          M+K+ALFAGGDL+Y   DFDYFVGIDRGSLL+NGL L+MAVAGDFDS+++   IK
10  Sbjct: 1  MSKVALFAGGDLSYISRDFDYFVGIDRGSLLLENGLPLNMAVAGDFDSVSQKAFTDIKEK 60

Query: 61  CSNIVSASAEKNDTDTTELALKTIFKEFPPEAQVTVFGAFGGRIDHMMSNIFLPSDRDLEPF 120
          ++A EKNDTDTTELALK +F FPEA+VT+FGAFGGR+DH++SNIFLPSD + PF
15  Sbjct: 61  AELFITAHPEKNDTDTTELALKEVFARFPEAEVTIFGAFGGRMDHLLSNIFLPSDPGIAPF 120

Query: 121  MSQIRLKDEQNIVTYLPSGKNQVSRIEGMSYVSFMPPESESTLQISGAKYELNKSNYFKKK 180
          M+QI L+D+QN++TY P+G++ + + EGM+YV+FM E E+ L I+GAK+EL + N+FKKK
20  Sbjct: 121  MAQIALRDQQNMITYRPAGQHLLHQEEGMYVAFMAEGEADLTITGAKFELTQDNFFKKK 180

Query: 181  MYSSNEFMTSPIEVELKDGYLIIIIYSKD 208
          +YSSN F+ PI V L GYLIII SKD
20  Sbjct: 181  IYSSNAFIHQPIITVSLPSGYLIIIIQSKD 208
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 1836

A DNA sequence (GBSx1943) was identified in *S.agalactiae* <SEQ ID 5701> which encodes the amino acid sequence <SEQ ID 5702>. This protein is predicted to be ribulose-phosphate 3-epimerase (rpe). Analysis of this protein sequence reveals the following:

30 Possible site: 18
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.59 Transmembrane 124 - 140 (124 - 141)

----- Final Results -----

```

35      bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB06221 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 113/211 (53%), Positives = 153/211 (71%)

```

Query: 5  KIAPSILAADYANFANELKRIEETTAEYVHIDIMDGQFVPNISFGADVSSMRKHSKLVF 64
          KIAPSIL+AD+AN NE++ +E A+Y+H+D+MDG FVPNI+ G +V ++R + L
45  Sbjct: 3  KIAPSILSADFANLGNEIQDVERGGADYIHVDVMDGHFVPNITIGPLIVDAIRPVTTLPL 62

Query: 65  DCHLMVVDPERYIEAFAQAGADIMTIHVEATKHIHGALQKIKEAGMKAGVVINPGTPVES 124
          D HLM+ P+ YI AFA+AGADI+T+HVEA H+H L IKE+G+KAGVV+NP TPV S
50  Sbjct: 63  DVHLMIEQPDGYIPAFKAGADIITVHVEACPHLHRTLHLIKESGVKAGVVLPATPVSS 122

Query: 125  LIPILDLVDQILIMTVNPGFGGQAFIPEMMSKVKTVAAWRKEYGHHYDIEVDGGIDNTTI 184
          + +L VD +L MTVNPGFGGQ FIP ++ K+K +A+ +KE G ++IEVDGG++ T
55  Sbjct: 123  IQHVLSDVDMVLFMTVNPGFGGQRFIPSVLPKLKELASLKKEQGLTFEIEVDGGVNEETA 182

Query: 185  KAAAEAGANVFVAGSYLKFASDLPAQVETLR 215
          K EAGANV VAGS +F D A ++ +R
60  Sbjct: 183  KQCVEAGANVLVAGSAVFNEEDRAAAIKGIR 213
  
```

-2069-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5703> which encodes the amino acid sequence <SEQ ID 5704>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0072(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 183/219 (83%), Positives = 198/219 (89%)

Query: 1 MSTNKIAPSILAADYANFANELKRIETTAEYVHIDIMDGGQFVPNISFGADVSSMRKHS 60

MST KIAPSILAADYANFA+EL RIEET AEYVHIDIMDGGQFVPNISFGADV+SMRKHS

Sbjct: 1 MSTLKIAPSILAADYANFASELARIEETDAEYVHIDIMDGGQFVPNISFGADVASM RKHS 60

Query: 61 KLVFDCHLMVVDPERYIEAFAQAGADIMTIHVEATKHIHGALQKIKEAGMKAGVVINPGT 120

KLVFDCHLMVVDPERY+EAFAQAGADIMTIH E+T+HIHGALQKIK AGMKAGVVINPGT

Sbjct: 61 KLVFDCHLMVVDPERYVEAFAQAGADIMTIHTESTRIHGALQKIKAAAGMKAGVVINPGT 120

Query: 121 PVESLIPILDLDVDQILIMTVNPGFGGQAFIPEMMSKVKTVAARKEYGHHYDIEVDGGID 180

P +L P+LDLVDQ+LIMTVNPGFGGQAFIPE + KV TVA WR E G +DIEVDGG+D

Sbjct: 121 PATALEPLLDLDVDQVLIMTVNPGFGGQAFIPECLEKVATVAKWRDEKGLSFDIEVDGGVD 180

Query: 181 NTTIKAAAEAGANVFVAGSYLKFASDLPAQVETLRVALD 219

N TI+A EAGANVFVAGSYLKFASDL +QV+TLR AL+

Sbjct: 181 NKTIRACYEAGANVFVAGSYLKFASDLVSQVQTLRTALN 219

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1837

A DNA sequence (GBSx1944) was identified in *S.agalactiae* <SEQ ID 5705> which encodes the amino acid sequence <SEQ ID 5706>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2098(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13451 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]

Identities = 148/296 (50%), Positives = 202/296 (68%), Gaps = 14/296 (4%)

Query: 2 QGRIVKSLAGFYV----ESDGVVYQTRARGNFRKKGQIPYVGDWVEFSSQDQSEGYILS 57

+G+I+K+L+GFYV E V Q R RG FRK P VGD+V + +++ EGY++

Sbjct: 3 EGKIIKALSGFYVLDSESDSKVIQCRGRGIFRKNKITPLVGDYVVYQAENDKEGYLME 62

Query: 58 IEERKNSLVRPPIVNIDQAVVIMSAKEPDFNANLLDRFLVLLEYKMIQPIIYISKLDLLD 117

I+ER N L+RPPI N+DQAV++ SA +P F+ LLDRFLVL+E IQPII I+K+DL++

Sbjct: 63 IKERTNELIRPPICNVDDQAVLVFSAVQPSFSTALLDRFLVLVEANDIQPIICITKMDLIE 122

Query: 118 DLVVIDDIR---EHYQNIGY-VFCYSQEE-----LLPLLANKVTVMGQTGVGKSTLLN 167

D D I+ E Y+NIGY V+ S ++ ++P +K TVF GQ+GVGKS+LLN

Sbjct: 123 DQDTEDTIQAYAEDYRNIGYDVYLTSSKDQDSLADIIPHFQDKTTFVAGQSGVGKSSLLN 182

Query: 168 KIAPELKLETGEISGLGRGHTTRAVSFYNVHKGIADTPGFSSLDYEVDNAEDLNESF 227

-2070-

I+PEL L T EIS LGRG+HTTR V + G +ADTPGFSSSL++ E+L +F
 Sbjct: 183 AISPELGLRTNEISEHLGRGKHTTRHVELIHTSGGLVADTPGFSSSLEFTDIEEEELGYTF 242

Query: 228 PELRRLSHFCKFRSCTHTHEPKCAVKEALTQQLWQVRYDNYLQFLSEIESRRETY 283
 P++R S CKFR C H EPKCAVK+A+ G+L Q RYD+Y++F++EI+ R+ Y
 Sbjct: 243 PDIREKSSSCKFRGCLHLKEPKCAVKQAVEDGELKQYRYDHYVEFMTEIKDRKPRY 298

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5707> which encodes the amino acid sequence <SEQ ID 5708>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2290(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 230/290 (79%), Positives = 257/290 (88%)

Query: 1 MQGRIVKSLAGFYVYESDGVVYQTRARGNFRKKGQIPYVGDWVEFSSQDQSEGYILSIEE 60
 +QG+I+KSLAGFYVYES+G VYQTRARGNFRK+G+ PYVGD V+FS++D SEGYIL+I
 Sbjct: 1 LQGKIIKSLAGFYVYESGQVYQTRARGNFRKRGETPYVGDIVDFSAEDNSEGYILAIHP 60

Query: 61 RKNSLVRPPIVNIDQAVVIMSAKEPDFNANLLDRFLVLEYYKMIQPIIYISKLDLLDDLIV 120
 RKNSLVRPPIVNIDQAVVIMSAKEP+FN+NLLDRFL+LLE+K I P++YISK+DLLD
 Sbjct: 61 RKNSLVRPPIVNIDQAVVIMSAKEPEFNSNLLDRFLILLEHKAIHPVYISKMDLLDSPE 120

Query: 121 VIDDIREHYQNIGYVFCYSQEELLPLANKVTVFMGQTGVGKSTLLNKIAPELKLETGEI 180
 I I YQ IGY F S EELLPLLA+K+TVFMGQTGVGKSTLLN+IAPEL LE GEI
 Sbjct: 121 EIKAIQRQYQAIGYDFVTSLEELLPLIADKITVFMGQTGVGKSTLLNRIAPELALEIGEI 180

Query: 181 SGSLGRGRHTTRAVSFYNVHKGIADTPGFSSLDYEDNAEDLNESFPFLRRLSHFCKFR 240
 S SLGRGRHTTRAVSFYN H GKIADTPGFSSLDY++ NAEDLNE+FPFLRRLSH CKFR
 Sbjct: 181 SDSLGRGRHTTRAVSFYNTHGGKIADTPGFSSLDYDIANAEDLNEAFPELRLSHECKFR 240

Query: 241 SCTHTHEPKCAVKEALTQQLWQVRYDNYLQFLSEIESRRETYKKVIKRK 290
 SCTHTHEPKCAVK AL G+LW VRY++YLQFLSEIE+RRETYKKVIKRK
 Sbjct: 241 SCTHTHEPKCAVKAALGTGELWPVRYEHYLQFLSEIENRRETYKKVIKRK 290

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1838

A DNA sequence (GBSx1945) was identified in *S.galactiae* <SEQ ID 5709> which encodes the amino acid sequence <SEQ ID 5710>. This protein is predicted to be rRNA. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.43 Transmembrane 259 - 275 (259 - 275)

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15937 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 95/278 (34%), Positives = 147/278 (52%), Gaps = 16/278 (5%)

-2071-

5 Query: 14 SYFACPKQNPLIKESN-SLKCDN-HCFDLSKFGYVNLGGKKVDEHYDKKSFENR-QL 70
S F CP C + + S SL C++ H FDLs+ GYVN L K V Y + FE R +L
Sbjct: 8 SMFRCPLCDSSMDAASGKSLICTERGHFTDLsRHGYVNFLT-KPVKTSYGAELEFEARSRL 66

10 Query: 71 VLENGYYNHILEAISKVLENNsqFH---SVLDIGCGEGFYSRQLVnkHEKtFLAF----D 123
+ E G+++ + +AI++++ + H ++LD GCGEG + L A D
Sbjct: 67 IGECGFFDPLHDAIAELISHPKSGHEAFTILDSGCGEGSHLNALCGFDYAGKAAIGTGID 126

15 Query: 184 VKVVPVAEHVQELREKASQYLKQKDYSNQKILDHFRENFEIIEQKVVQSYNCSQQERQA 243
+KVVP ++++ ELR+ ++ YSN ++ F N ++ QQ
Sbjct: 185 IKVVRSDYLIELRQFLYTDSPrTYSNTAAVERFTANAHSRPVRLRYVKTLDDQQAIIHW 244

20 Query: 244 FIDMTPLLFsVDKTTIDW---ASISEITVGALIVIGKK 278
+ MTPL +S K + ++ITV I+IG K
Sbjct: 245 LLKMTPLAWSAPKDRVSLLEKMSADITVDVILIGMK 282

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 1839

A DNA sequence (GBSx1946) was identified in *S.agalactiae* <SEQ ID 5711> which encodes the amino acid sequence <SEQ ID 5712>. This protein is predicted to be dimethyladenosine transferase (ksgA). Analysis of this protein sequence reveals the following:

30 Possible site: 61
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3257(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB11818 GB:Z99104 dimethyladenosine transferase [Bacillus subtilis]
Identities = 157/284 (55%), Positives = 215/284 (75%), Gaps = 2/284 (0%)

45 Query: 3 IADKTVTRAILERHGFTFKKSFGQNFLTDTNlQKIVDTAEIDKGVNVVIEIGPGIGALTE 62
IA T+ IL+++GF+FKKS QQNFL DTNl +IVD AE+ + VIEIGPGIGALTE
Sbjct: 5 IATPIRTKEILKKYGFSFKKSLGQNFLIDTNlNRIVDHAEVTEKTGVIEIGPGIGALTE 64

50 Query: 63 FLAENAAEVMAFEIDDRILIPILADTLARFDNVQVNVQDILKADLQTQIQa-FKNPDLPIK 121
LA+ A +V+AFEID RL+PIL DTL+ ++NV V++QD+LKAD+++ I+ F++ D I
Sbjct: 65 QLAKRAKKVVAFEIDQRLLPILKDTLSPYENVTVIHQDVLKADVKSVEEQFQDCD-EIM 123

55 Query: 122 VVANLPYYITTPILMHLIESKIPFAEFVVMIOKEVADRISAMPNTKAYGSLSIaVQYYMT 181
VVANLPYY+TTPI+M L+E +P VVM+QKEVA+R++A P++K YGSLSIaVQ+Y
Sbjct: 124 VVANLPYYVTTPIMKLLEHLPLKGIVVMLQKEVAERMAADPSKEYGSLSIaVQFYTE 183

60 Query: 182 AKVSFIVPRTVFVPAPNVDSAILKMVRDQPVVSVQDEDFFRVSKVAFVHRRKTLWNNL 241
AK IVP+TVFVP PNVDSA++++ RD P V V++E FFF++ K +F RRKTL NNL
Sbjct: 184 AKTVMIVPKTVFVPQPNVDSAVIRLILRDGPAVDVENESFFFQLIKASFAQRRKTLNLL 243

Query: 242 TSHFGKSEDTKAKLEKALEIAKIKPSIRGEALSIPDFASLADAL 285
++ + + K+ +E+ LE I RGE+LSI +FA+L++ L
Sbjct: 244 VNNLPEGKAQKSTIEQVLEETNIDGKRRESLSTEEFAALSNGL 287

-2072-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5713> which encodes the amino acid sequence <SEQ ID 5714>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2420(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 257/290 (88%), Positives = 275/290 (94%)

Query: 1 MRIADKTVTRAILERHGFTEFKKSFGQNFLTDTNQLKIVDTAEIDKGVNVIEIGPGIGAL 60

MRIAD +VT+A+L+RHGFTEFKKSFGQNFLTDTNQLKIVDTAEID+ VNVIEIGPGIGAL

Sbjct: 9 MRIADYSVTKAVLDRHGFTEFKKSFGQNFLTDTNQLKIVDTAEIDQNVNVIEIGPGIGAL 68

Query: 61 TEFLAENAAEVMMAFEIDRLPILADTLARFDNVQVVDILKADLQTQIQAFKNPDLPI 120

TEFLAENAAEVMMAFEIDRL+PILADTL FDNVQVVDILKADLQTQI+ FKNPDLPI

Sbjct: 69 TEFLAENAAEVMMAFEIDRLVLPILADTLRDFDNVQVVDILKADLQTQIKQFKNPDLPI 128

Query: 121 KVVANLPYYITTPILMHLIESKIPFAEFVVMQKEVADRISAMPNTKAYGSLSIQVQYYM 180

KVVANLPYYITTPILMHLIESKIPF EFVVM+Q+EVADRISA PNTKAYGSLSIQVQYYM

Sbjct: 129 KVVANLPYYITTPILMHLIESKIPFQEFVVMQREVADRISAEPNTKAYGSLSIQVQYYM 188

Query: 181 TAKVSFIVPRTVFPAPNVDSAILKMMVRDQPVSVQDEDFFRVSKVAFVHRRKTLWNN 240

TAKV+FIVPRTVFPAPNVDSAILKMMVRDQPV+ V+DEDFFRVSKVAFVHRRKTLWNN

Sbjct: 189 TAKVAFIVPRTVFPAPNVDSAILKMMVRDQPLIEVKDEDFFRVSKVAFVHRRKTLWNN 248

Query: 241 LTSHFGKSEDTKAKLEKALEIAKIKPSIRGEALSIPDFASLADALKEVGI 290

LTSHFGKSED KAKLEK L +A IKPSIRGEALSIP DF LADALKEVG+

Sbjct: 249 LTSHFGKSEDIKAKLEKGLALADIKPSIRGEALSIPDFGKLADALKEVGL 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1840

A DNA sequence (GBSx1947) was identified in *S.agalactiae* <SEQ ID 5715> which encodes the amino acid sequence <SEQ ID 5716>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0736(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1841

A DNA sequence (GBSx1948) was identified in *S.agalactiae* <SEQ ID 5717> which encodes the amino acid sequence <SEQ ID 5718>. Analysis of this protein sequence reveals the following:

-2073-

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3031(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB11817 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 81/179 (45%), Positives = 117/179 (65%), Gaps = 4/179 (2%)

Query: 7 IQEVIVVEGKDDTANLRRFYNVDTYETRGS AIDEDDLERLHNLRGVIVFTDPDYNGE 66
 I+E+IVVEG+DDTA ++ + DT ET GSAID+ +++I RGVI+ TDPD+ GE
 15 Sbjct: 3 IKEIIVVEGRDDTARIKLAVDADTIETNGSAIDHVIDQIRLAQKTRGVIIITDPDFPGE 62

Query: 67 RIRKIIMNAIPTVRHAFNLNDEAKPGSKTKGRSLGVEHASFEDLQKALSKVTQHFDDEDH 126
 +IRK I A+P +HAFL + AKP +K R +GVEHAS E ++ L V + + +
 20 Sbjct: 63 KIRKTISEAVPGCKHAFLPKHLAKPKNK---RGIGVEHASVESIRACLENVHEEMEAQPS 119

Query: 127 FDITQADLIRWGFITASDSRKRREYLG NQLRIGYSNGKQLLKRRLFGVTKAEVEECME 185
 DI+ DLI G I ++ RRE LG+ L+IGY+NGKQL KRL++F + K++ ++
 25 Sbjct: 120 -DISAEDLIHAGLIGGPAAKCRRERLGDLLKIGYTNGKQLKRLQMFQIKKSDFMSALD 177

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5719> which encodes the amino acid
 sequence <SEQ ID 5720>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1474(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 146/187 (78%), Positives = 165/187 (88%)

Query: 1 MMKKIDIQEVIVVEGKDDTANLRRFYNVDTYETRGS AIDEDDLERLHNLRGVIVFTD 60
 + +KI+IQEV+VVEGKDDTANLRRFY VDTYETRGS A I E+DLERI RL++LRGVIV TD
 40 Sbjct: 15 LTEKINIQEVIVVEGKDDTANLRRFYVDTYETRGS A ITEEDLERINRLNDLRGVIVLTD 74

Query: 61 PDYNGERIRKIIMNAIPTVRHAFNLNDEAKPGSKTKGRSLGVEHASFEDLQKALSKVTQH 120
 PDYNGERIRK+IM A+PT RHAFLNR+EA P SK+KGRSLGVEHA+FEDLQKAL+ VTQ
 45 Sbjct: 75 PDYNGERIRKLIMAAVPTARHAFNLNDEAVPSSKSKGRSLGVEHANFEDLQKALAHVTQQ 134

Query: 121 FDEDEHFDITQADLIRWGFITASDSRKRREYLG NQLRIGYSNGKQLLKRRLFGVTKAEV 180
 +DDE +FDI Q DLIR G + ASDSRKRREYLG +LRIGY+NGKQLLKRRL LFG+T AEV
 50 Sbjct: 135 YDDSYFDIRQTDLIRLGLIMASDSRKRREYLG EKLIRIGYANGKQLLKRLELFGITLAEV 194

Query: 181 EECMGY 187
 EE ME Y
 50 Sbjct: 195 EEVMET 201

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 55 vaccines or diagnostics.

Example 1842

A DNA sequence (GBSx1949) was identified in *S.agalactiae* <SEQ ID 5721> which encodes the amino
 acid sequence <SEQ ID 5722>. Analysis of this protein sequence reveals the following:

Possible site: 15

-2074-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4955(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10139> which encodes amino acid sequence <SEQ ID 10140> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11815 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 115/254 (45%), Positives = 172/254 (67%)

15 Query: 28 IFDTHTHLNVENFEGKIDEEINLASELGVTKMNVVGFQDTISKSLSSQYQVYSTIG 87
 +FDTH HLN E ++ ++E I A V ++ VVGFD+ TI++++E+ +Y +Y+ IG
 Sbjct: 2 LFDTHAHLNAEQYDLDLEEVIERAKAEKVERIVVVGFDRTITRAMEMIEEYDFIYAAIG 61

20 Query: 88 WHPTEAGSYDDNIESMIISHLENPKVIALGEIGLDYYWMEDPKDIQIEVFKRQIELSKEY 147
 WHP +A + + I + KV+A+GE+GLDY+W + PKDIQ EVF+ QI L+KE
 Sbjct: 62 WHPVDAIDMTEDDLAWIKELSAHEKVVAIGEMGLDYHWDKSPKDIQKEVFRNQIALAKEV 121

25 Query: 148 NLPFVVHTRDALEDITYEVIKESGVGPFPGGIMHSFSGSLEMAQKFIDLGMMSFSGVVTFFK 207
 NLP ++H RDA ED ++KE G GGIMH F+GS E+A++ + + +SF G VTFK
 Sbjct: 122 NLPIIIHNRDATEDVVTILKEEGAEAVGGIMHCFTGSAEVARECMKNFYLSFGGPVTFK 181

30 Query: 208 KALDVQEAARELPLDKILVETDAPYLAPVPKRGRENKTAYTRYVVEKIAELRGITVEEVA 267
 A +E +E+P D++L+ETD P+L P P RG+ N+ +Y +YV E+IAEL+ +T EE+A
 Sbjct: 182 NAKKPKKEVVKIEPNDRLLIETDCPFLTPHPFRGKRNEPSYVKYVAEQIAELKEMTFEEIA 241

30 Query: 268 EATYQNAVRIFRLD 281
 T +NA R+FR++
 Sbjct: 242 SITTENAKRLFRIN 255

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5723> which encodes the amino acid sequence <SEQ ID 5724>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 190/258 (73%), Positives = 227/258 (87%)

50 Query: 24 DMIKIFDTHTHLNVENFEGKIDEEINLASELGVTKMNVVGFQDTISKSLSSQYQVY 83
 + + IFDTHTHLVN F+G EE+ LA E+GV NVVGFQQ TIS +L L+++YA +Y
 Sbjct: 38 EKLTI FDTHTHLNVAEFQGHETEELTLAQEMGVAYHNVVGFQQTATISGALT LANKYANIY 97

55 Query: 84 STIGWHPTEAGSYDDNIESMIISHLENPKVIALGEIGLDYYWMEDPKDIQIEVFKRQIEL 143
 +TIGWHPTEAGSY + +E I+S L + KVIALGEIGLDYYWMEDPK++QIEVFKRQ++L
 Sbjct: 98 ATIGWHPTEAGSYSEAVEEAIVSQLSHSKVIALGEIGLDYYWMEDPKQIEVFKRQMQQL 157

60 Query: 144 SKEYNLFPVVHTRDALEDITYEVIKESGVGPFPGGIMHSFSGSLEMAQKFIDLGMMSFSGV 203
 +K+++LFPVVHTRDALEDITYEVIK +GVGP GGIMHS+SGSLEMA++FI+LGMMISFSGV
 Sbjct: 158 AKDHDLPFVVHTRDALEDITYEVIKAAAGVGPRGGIMHSYSGSLEMAERFIELGMMSFSGV 217

60 Query: 204 VTFKKALDVQEAARELPLDKILVETDAPYLAPVPKRGRENKTAYTRYVVEKIAELRGITV 263
 VTFKKALD+QEAA+ LPLDKILVETDAPYL VVPKRG++N TAYTRYVV+KIAELRG+TV
 Sbjct: 218 VTFKKALDIQEAAQHLPLDKILVETDAPYLTVPKRGKQNHNTAYTRYVVDKIAELRGMTV 277

-2075-

Query: 264 EEVAEATYQNAVRIERLD 281
 EEVA+AT NA R+F+LD
 Sbjct: 278 EEVAKATTANAKRVFKLD 295

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1843

- A DNA sequence (GBSx1950) was identified in *S.agalactiae* <SEQ ID 5725> which encodes the amino acid sequence <SEQ ID 5726>. This protein is predicted to be endosome-associated protein. Analysis of
 10 this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5142(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1844

- A DNA sequence (GBSx1951) was identified in *S.agalactiae* <SEQ ID 5727> which encodes the amino acid sequence <SEQ ID 5728>. This protein is predicted to be CG17785 gene product. Analysis of this
 25 protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

- 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4730(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1845

- 40 A DNA sequence (GBSx1952) was identified in *S.agalactiae* <SEQ ID 5729> which encodes the amino acid sequence <SEQ ID 5730>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4032(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2076-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:BAB01041 GB:AB022220 gene_id:MLN21.14~unknown protein
    [Arabidopsis thaliana]
    Identities = 49/185 (26%), Positives = 85/185 (45%), Gaps = 46/185 (24%)

Query: 5  LTDLDRVNIKQEYELGSQLDTLVKIMSQDKVLPIGKVAHVQ-----DGGKETGEQIYT 58
          L +D V+ + + ELGS+ + +M+ K+ V+ D K+ Q++
10  Sbjct: 154 LEGIDSVDSCRVKIELGSRGLMDLCVMASKLAYENAKMNLVEFLDCWNDYQKMSTQVFV 213

Query: 59  ITPNGTLDKPEDVKEVTVLFGKSTAPFGGDDWKTID---WFKNDIPIASKL---LLKKFG 111
          T DK +D + + F+G T PF DDW TD W+ ++P KL L+ G
15  Sbjct: 214 FT-----DKQKDANLIVISFRG-TEPFDADDWGTDYDFYSWY--EVPNVGKLHMGFLEAMG 265

Query: 112 -----SQSVSHKQGTKQ-----LEQSAH-----LLKEVMNKYPNAKISVY 146
          Q+ S ++ +K+ +E+SA+ +LK +++++ NA+ V
15  Sbjct: 266 LGNRDDT'TTFHYNLFEQTSSEEENSKKNLLDMVERSAYYAVRVILKRLLEHENARFVVT 325

20  Query: 147 GHSLG 151
          GHSLG
    Sbjct: 326' GHSLG 330

```

No corresponding DNA sequence was identified in *S.pyogenes*.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1846

A DNA sequence (GBSx1953) was identified in *S.agalactiae* <SEQ ID 5731> which encodes the amino acid sequence <SEQ ID 5732>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 52
    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -8.97    Transmembrane    12 - 28 ( 5 - 33)

    ----- Final Results -----
35  bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 A related GBS nucleic acid sequence <SEQ ID 10141> which encodes amino acid sequence <SEQ ID 10142> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45 A related GBS gene <SEQ ID 8909> and protein <SEQ ID 8910> were also identified. Analysis of this protein sequence reveals the following:

```

50  Lipop: Possible site: -1    Crend: 4
    McG: Discrim Score:    14.01
    GvH: Signal Score (-7.5): -5.55
    Possible site: 46
    >>> Seems to have an uncleavable N-term signal seq
    ALOM program    count: 1 value: -8.97 threshold: 0.0
    INTEGRAL    Likelihood = -8.97    Transmembrane    6 - 22 ( 1 - 27)

```

-2077-

PERIPHERAL Likelihood = 9.49 84
modified ALOM score: 2.29

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 SEQ ID 8910 (GBS32) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 10 (lane 2; MW 15.6kDa).

GBS32-His was purified as shown in Figure 191, lane 8.

Example 1847

15 A DNA sequence (GBSx1954) was identified in *S.agalactiae* <SEQ ID 5733> which encodes the amino acid sequence <SEQ ID 5734>. This protein is predicted to be extramembranal protein (dltD). Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -10.24 Transmembrane 12 - 28 (4 - 31)

----- Final Results -----

bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC29041 GB:AF050517 unknown [Streptococcus mutans]

Identities = 242/421 (57%), Positives = 309/421 (72%), Gaps = 1/421 (0%)

30 Query: 1 MLKRLGKVFGLVLCALLLLVGLYFVFPVSQ- PHLGKEKNSAVALTKAGFKSRVQKVRAF 59
MLKRL + GP+ CAL+L+ L +P H+ +EKN AVAL+ + FKS +K+RA
Sbjct: 1 MLKRLWLILGPVFCALVLFSLIMFYPAKHLSHNYNEEKNDVALSPSSFKSTNKKMRAL 60

35 Query: 60 SDPKANFVPPFGSSEWLRFDMHPSVLAERNSYIPYLLGQKGAASLTQYYGIQQIKGQ 119
SD + FVPPFGSSEW R D MHPSVLAERNSY PYLLGQKG+ SL+ Y+G+QQI Q
Sbjct: 61 SDKRHLFVPPFGSSEWQRIDNMHPSVLAERYNSYRPYLLGQKGSTSLSHYFGMQQIGNQ 120

40 Query: 120 IKNKKAIFYVISQWFVRKGANKGAFQNYFSNDQIRFLQNTGTTYDRYAARRLLKLYPE 179
IKNKKA+YVISQWFV KG + AFQ YFS++Q FL NQTG+T DRYAA+RLL + P
Sbjct: 121 IKNKKAVYVISQWFVPKGTSPIAFQYFSSEQLADFLNQTGSTADRYAAKRLLDIKPS 180

45 Query: 180 ASMSDLIEKVADGQKLSNKDKQRLKFNDWVFEKTD AIFSYP LPLGKTYNQAIMPHVKGKLPK 239
+++ +I+K+A G+ L++ D+ L+ +K DA+F L Y + ++PHV KLPK
Sbjct: 181 SNLQGMIIKIAAGKTLNSFDRASLRLIKSFLLKEDALFGSLTFSDNYERRVLPKLVKLPK 240

Query: 240 AFSYNHLSRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKLLKGSQRHFNYYTKSPEFNDL 299
FSY LS+IAS+D + T++NQF I+D FY RIK LK+LKG Q+ +Y +SPE+NDL
Sbjct: 241 HFSYGTLSQIASKDQRLTKTNQFEINDHFYNKRIKQQLKRLKGFQKQLSYLQSPEYNDL 300

50 Query: 300 QLVLNEFSKQNTDVLVFI PPVNNKKTIDYTGLDQMYQKSVEKIKHQLSQGFNHIADLSR 359
QL L + +K T V+FVIPPVN KW +YTGL Q MYQK+VEKIK+QLQSQGF++IADLS+
Sbjct: 301 QLALTQLAKSKTKVIFVIPPVNAKWVEYTGLSQDMYQKTVEKIKYQLQSQGFNDIADLSK 360

55 Query: 360 DGGKPYFMQDTIHLGWNGWLELDKHINPFLTEENSKPNYHINNKLKSWAKYTGRRPSDYK 420
+G +PYFMQDTIHLGWNGWL DK +NPFL+++ +P Y INN FL K WA YTG P +K
Sbjct: 361 NGDQPYFMQDTIHLGWNGWLAFDKEVNPFLSKKQLPAYKINNHLKSKWATYTGPNPFQFK 421

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5735> which encodes the amino acid sequence <SEQ ID 5736>. Analysis of this protein sequence reveals the following:

-2078-

Possible site: 41

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -13.06 Transmembrane 7 - 23 (1 - 31)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.6222(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 209/410 (50%), Positives = 278/410 (66%)

Query: 1 MLKRLGKVFGLVLCALLLLVGLYFVFPVSPHHLGKEKNSAVALTKAGFKSRVQKVRAFS 60
MLKRL + GPL+ A +L+V F FP H + +EK +AVA+T + FK+ + K +A S

15 Sbjct: 1 MLKRLWLILGPLLIAFVLVITIFSFPQTQLDHSIAQEKANAVAITDSSFKNGLIKQALS 60

Query: 61 DPKANFVPPFGSSEWLRFDMHPSVLA EAYNRSYIPYLLGQKGAASLTQYYGIQIQIKQOI 120
D FVPPFGSSEW R D+MHPSVLA E Y RSY P+L+G++G+ASL+ YYGIQOI ++

20 Sbjct: 61 DETCRFVPPFGSSEWSRMDSMHPSVLAERYKRSYRPFLLIGKGSASLSHYIYGIQITNEM 120

Query: 121 KNKKAIYVISPOWFVRKGANKGAFQNYFSNDQTIRFLQNTGTTYDRIAARRLLKLYPEA 180
+ KKAI+V+SPQWF +G N A Q Y SN Q I FL ++AA+RLL+L P

Sbjct: 121 QKKKAIFFVSPQWFTAQGINPSAVQMYLSNTQVIEFLKARTDKESQFAAKRLLLELNPV 180

Query: 181 SMSDLIEKVADGQKLSNKKQRLKFNDWVFEKTD AIFSYLPLGKTYNQAIMPHVGKLPKA 240
S S+L++KV+ G+ LS D+ LK V + +++FS+L Y + I+P V LPK

25 Sbjct: 181 SKSNLLKKVSKGKSLSRDLRAILKQCQHQVALREESLFSFLGKSTNYEKRI LPRVKG LPKV 240

Query: 241 FSYNHL SRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKGSRHFN YTKSPEFNDLQ 300
FSY L+ +A++ ++AT +N+FGI + FY+ RI K Q +++Y SPE+ND Q

30 Sbjct: 241 FSYKQLNALATKRGQLATNNRFGIKNTFYRKRIAPKYNLYKNFQVNYSLASPEYND FQ 300

Query: 301 LVLNEFSKQNTDVL FVIPPVNKKWTDYTG LDKMYQKSVEKIKHQLQSQGFNHIADLSRD 360
L+L+EF+K+ TDVLFVI PVNK W DYTGL+Q YQ +V KIK QL+SQGF+ IAD S+D

35 Sbjct: 301 LLLSEFAKRKTDVLFVITPVNKAWADYTGLNQDKYQAAVRKIKFQLKSQGFHRIADFSKD 360

Query: 361 GSKPYFMQDTIHLGWNGWLELDKHINPFLTEENSKPNYHINNKFLLKSWA 410

GG+ YFMQDTIHLGWNGWL DK + PFL + PNY +N F K WA

Sbjct: 361 GGESYFMQDTIHLGWNGWLA FDKKVPFLETQKQVPVNYKMNPFYFSKIWA 410

40

A related GBS gene <SEQ ID 8911> and protein <SEQ ID 8912> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6

McG: Discrim Score: 15.50

45 GvH: Signal Score (-7.5): -4.52

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 1 value: -10.24 threshold: 0.0

INTEGRAL Likelihood = -10.24 Transmembrane 12 - 28 (4 - 31)

50 PERIPHERAL Likelihood = 8.33 301

modified ALOM score: 2.55

*** Reasoning Step: 3

55 ----- Final Results -----

bacterial membrane --- Certainty=0.5097(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

57.5/76.3% over 420aa

Streptococcus mutans

GP|3403204| unknown Insert characterized

-2079-

ORF00336(301 - 1560 of 1860)
 GP|3403204|gb|AAC29041.1||AF050517(1 - 421 of 421) unknown {Streptococcus mutans}
 %Match = 41.0
 %Identity = 57.5 %Similarity = 76.2
 Matches = 242 Mismatches = 99 Conservative Sub.s = 79

```

33      63      93      123      153      183      213      243
FSGFLDLLWFPQPHNK**GVL*WILNQKY*QLLMTYLWRMELL*WMKTYLTQEF*TAWVLLN*LLSWKATLILIFRLRLNL

10  273      303      333      363      420      450      480
VVMGTGTLIKLLE*RSSAMLKRLGKVFGLVLCALLLVGLYFVFPVSQ-PHHLGKEKNSAVALTKAGFKSRVQKVRAF$
      ||||| :||: |||:| :| :| :| :|| ||||: : ||| :||:|
      MLKRLWLILGPVFCALVLVFSLIMFYPAKHL$HNYNEEKND$AVALSP$SFKSTNKKMRALS
      10      20      30      40      50      60

15  510      540      570      600      630      660      690      720
DPKANFVPPFFGSSEWLRFDAMHPSVLA$EAYNRSYIPYLLGQKGAASLTQYYGIQKQIKNKKAIYVISPQW$FVRKGAN
| : ||||| ||| ||||| ||||| ||||| ||||| :||:| :||:| ||||| :||| ||| :||:|
DKRHLFVPPFFGSSEWQRIDNMHPSVLA$ERYNRSYRYPYLLGQKGSTSLSHYFGMQQIGNQIKNKKAVYVISPQW$FVPGKTS
20      80      90      100      110      120      130      140

750      780      810      840      870      900      930      960
KGAFQNYFSNDQTIRFLQNQTGTTYDRYA$ARRLLKLYPEASMSDLIEKVADGQKLSNKKQRLKFNDWVFEKTD$AIFSYL
||| ||||:| || ||||:| ||||| :|| :| :||:| ||| :||:| :|| :||:|
PIAFQQYFSSEQLADFLNQTGSTADRYA$AKRLLDIKPSSNLQGMIKKIAAGKTLNSFDRASLRLIK$SFLKKEDALFGSL
25      160      170      180      190      200      210      220

990      1020      1050      1080      1110      1140      1170      1200
PLGKTYNQAIMPHV$GKLPKAFSYNHL$SRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKG$SRHFNYTK$SPEFNDLQ
: : :||| |||| ||| ||||:| :| :||| :| || ||| :||:| ||| :||:| :||:|
TFSDNYERRVLP$HVKKLPKHFSYGTLSQIASKDQGRLLTKTNQFEINDH$FYNKRIKGQLKRLKGFQKLSY$LSQ$PEYNDLQ
30      240      250      260      270      280      290      300

1230      1260      1290      1320      1350      1380      1410      1440
LVLNEFSKQNTDVL$FVIPPV$NKKWTDYTGLDQKMYQKSVEKIKHQLQSQGFNHIADLSRDG$GKPYFMQDTIHL$GWN$WLE
| :||:| | :||| ||| :||| | ||||:| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
LALTQ$AK$TKVIFVIPPV$NAKWVEYTG$LSQDMYQKTVEKIKYQLQSQGF$DNIADLSKNGDQPYFMQDTIHL$GWN$WLA
35      320      330      340      350      360      370      380

1470      1500      1530      1560      1590      1620      1650      1680
LDKHINPFLTEEN$SKPNYHINNKFLK$KSWAKYTG$RPSDYK*IVESDDL*H*$SY*$SFLISLYLVILR*LIHVL*$FFIYNE
:| :||| :||:| :| ||| ||| ||| ||| :|
FDKEVNPFL$KQ$LPAYKINN$HFL$K$KWATYTG$NPFQ$FK
40      400      410      420

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1848

A DNA sequence (GBSx1955) was identified in *S.agalactiae* <SEQ ID 5737> which encodes the amino acid sequence <SEQ ID 5738>. This protein is predicted to be d-alanyl carrier protein (dltC). Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.1061(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05776 GB:AF051356 D-alanyl carrier protein [Streptococcus mutans]

-2080-

Identities = 65/79 (82%), Positives = 74/79 (93%)

Query: 1 MDIKSEVLAIIDDLFMEDVSSMMDEDLFDAGVLDMSGTVLIVELESHFNIDIPAEFGR 60
 MDIKSEVL IID+LFMEDVS MMDEDLFDAGVLDMSGTVLIVELE+HF+I +P++EFGR
 5 Sbjct: 1 MDIKSEVLKIIDELEFMEDVSDMMDEDLFDAGVLDMSGTVLIVELEHNFIDITVPVSEFGR 60

Query: 61 NDWNTANKIVAGVTELCNA 79
 +DWNTANKI+ G+TEL NA
 10 Sbjct: 61 DDWNTANKIIEGITELRNA 79

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5739> which encodes the amino acid sequence <SEQ ID 5740>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3976(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 57/79 (72%), Positives = 65/79 (82%)

Query: 1 MDIKSEVLAIIDDLFMEDVSSMMDEDLFDAGVLDMSGTVLIVELESHFNIDIPAEFGR 60
 M I+ V+ + D LFMEDVS MMDEDLFDAGVLD+GTVELIVELES FNI +PI+EFGR
 25 Sbjct: 1 MSIEETVIELEFDRLEFMEDVSEMMDEDLFDAGVLD+SLGTVELIVELESTFNIKVPISFGR 60

Query: 61 NDWNTANKIVAGVTELCNA 79
 +DWNT KIV GV EL +A
 30 Sbjct: 61 DDWNTVTKIVQGVVEELQHA 79

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1849

A DNA sequence (GBSx1956) was identified in *S.agalactiae* <SEQ ID 5741> which encodes the amino acid sequence <SEQ ID 5742>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.55 | Transmembrane | 93 - 109 (91 - 117) |
| INTEGRAL | Likelihood = -7.64 | Transmembrane | 21 - 37 (19 - 39) |
| INTEGRAL | Likelihood = -6.79 | Transmembrane | 390 - 406 (387 - 410) |
| INTEGRAL | Likelihood = -5.20 | Transmembrane | 41 - 57 (40 - 59) |
| INTEGRAL | Likelihood = -2.07 | Transmembrane | 203 - 219 (200 - 221) |
| INTEGRAL | Likelihood = -1.65 | Transmembrane | 65 - 81 (65 - 81) |
| INTEGRAL | Likelihood = -0.75 | Transmembrane | 125 - 141 (125 - 141) |

----- Final Results -----

bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5743> which encodes the amino acid sequence <SEQ ID 5744>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have an uncleavable N-term signal seq

| | | | |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.14 | Transmembrane | 387 - 403 (382 - 409) |
| INTEGRAL | Likelihood = -9.66 | Transmembrane | 18 - 34 (15 - 37) |
| INTEGRAL | Likelihood = -5.95 | Transmembrane | 64 - 80 (63 - 81) |

-2081-

INTEGRAL Likelihood = -5.63 Transmembrane 92 - 108 (89 - 114)
 INTEGRAL Likelihood = -1.97 Transmembrane 40 - 56 (40 - 56)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAC05775 GB:AF051356 integral membrane protein [Streptococcus mutans]
 Identities = 246/413 (59%), Positives = 319/413 (76%)

Query: 1 MMMFFSHIPYMEPYGNPIYFVYLILAFPLVPIIGIFKQKRLSTYETLVSLVFIPLFMFGGDH 60
 M+ FF ++P++E YGNP YF Y+ILA LP+ IG+F +KR YE VSL+FI+ M G+

15 Sbjct: 1 MIDFFKNLPHLEAYGNPQYFFYIILAVLPIFIGLFFKKRFPLYEAFVSLIFIVMLTGEK 60

Query: 61 YQQLVAFLLFYLLWQIISVFAYQKYRENANSAGVFYLAIAMALFPLIWKVAPLTGPSSQT 120
 Q+ A FY++WQI V++Y+ YR++ ++ +FYL + M++ PL VK+ P + Q+

20 Sbjct: 61 SHQIFALFFYIIWQIFCVYSYKFKRSRDNKWIFYLHVMSILPLSLVKITPAIWTNQQS 120

Query: 121 LFSFLGISYLTFSKISGMIIEMRDGTLQEVRLPDFIRFMIFPPTFSSGPIDRFRRHFQEDYH 180
 LF FLGISYLTFS+S+GMI+EMRDG L +FIRFM+F PTFSSGPIDRF R +DY

25 Sbjct: 121 LFGFLGISYLTFRSVGMIMEMRDGVLTSFTFWFIRFMLFMPTFSSGPIDRFRRFNDDYE 180

Query: 181 KLPERDDYFAMLNKAVMYMLMGFLYKHIISYCLGGILLPLENKMVMGGYFNKETILVM 240
 K+P++D+ ML ++V Y+MLGF YK +++ LG ++LP L+ AL GG+FN T+ VM

30 Sbjct: 181 KIPDKDELDMLEQSVHYIMLGFFYKFVLAQILGTMLPGLKEMALQKGGWFNWPITLGV 240

Query: 241 YVYGLNLFFDFAGYSMFAIGISYLLGIRTPENFNMPFLSASLKDFWNRWHMSLSFWFRDY 300
 YVYGL+LFFDFAGYSMFAI IS +GI++P NFN PF S LK+FWRWHMSLSFWFRD+

35 Sbjct: 241 YVYGLDLFFDFAGYSMFAIAISNFMGIKSPTNFNQPFKSQDLKEFWNRWHMSLSFWFRDF 300

Query: 301 VFMRVLVHLLIKHKTFKNNRNVTSVAYLVNMLVMGFHWGLTWYYIAYGLFHGIGLIINDAW 360
 VFMRLV +L+K+K FKNRNVTS VAY+VNML+MGFWHG+TWYYI YGLFHG+GL++NDAW

40 Sbjct: 301 VFMRVLVKLVKNKVFKNRNVTSVAYIVNMLIMGFHWGVTWYYIITYGLFHGVGLVINDAW 360

Query: 361 IRKKKEINRHRKKKGLSPLFQSRAFHVLCIVVTFHVVMFSLLLFSGFLNDLWF 413
 +RKKK +N+ RK K LSPL ++ L IV+TF+VVM S L+FSGFLNDLWF

45 Sbjct: 361 LRKKKRLNKKERKAKNLSPLPENGWTRALGIVITFNVVMSLFLFSGFLNDLWF 413

An alignment of the GAS and GBS proteins is shown below.

Identities = 240/416 (57%), Positives = 317/416 (75%), Gaps = 5/416 (1%)

45 Query: 5 FLEKPLHLDVYGNPQYFFYLILAVLPIYIGLFFKKRFALYEIIFSLSFIVMMLTGSTFNQ 64
 F +P+++ YGNP YF YLILA LP+ IG+F +KR + YE + SL FI+ M G + Q

Sbjct: 4 FFFSHIPYMEPYGNPIYFVYLILAFPLVPIIGIFKQKRLSTYETLVSLVFIPLFMFGGDHYQQ 63

Query: 65 LKSLAYVVGQSLLVFIYKAYKRNFNHTLVFYVTVCLSIFPLFLVKLIPAISEDGHQSLF 124
 L + L Y++ Q + VF Y+ YR+ N VFY+ + ++FPL VK+ P ++ Q+LF

50 Sbjct: 64 LVAFLFYLLWQIISVFAYQKYRENANSAGVFYLAIAMALFPLIWKVAP-LTGPSSQTLF 122

Query: 125 GFLGISYLTFRVAMIIEMRDGVLKEFTLWBFLLFFPTFSSGPIDRFKRFNEDYINI 184
 FLGISYLTFR+++ MIIEMRDG L+E L +F+RF++FFPTFSSGPIDRF+ F EDY +

55 Sbjct: 123 SFLGISYLTFSKISGMIIEMRDGTLQEVRLPDFIRFMIFPPTFSSGPIDRFRRHFQEDYHKL 182

Query: 185 PDRNELLDMLGQAIHYLMGLFYKFIAYIFGSLIMPPLKELALEQGGVFNWPTLGVMYA 244
 P+R++ ML +A+ YLMGLFYK I++Y G +++P L+ AL GG FN T+ VMY

60 Sbjct: 183 PERDDYFAMLNKAVMYMLMGFLYKHIISYCLGGILLPLENKMVMGGYFNKETILVMYV 242

Query: 245 FGFDLFFDFAGYTMFALAI SNLMGIKSPINFDPKPKSRDLKEFWNRWHMSLSFWFRDFVF 304
 +G +LFFDFAGY+MFA+ IS L+GI++P NF+ PF S LK+FWRWHMSLSFWFRD+VF

65 Sbjct: 243 YGLNLFFDFAGYSMFAIGISYLLGIRTPENFNMPFLSASLKDFWNRWHMSLSFWFRDYVF 302

Query: 305 MRLVKLLVKNKVFKNRNVTSVAYIINMLMGFWHGLTWYYIAYGLFHGIGLVINDAWVR 364
 MRLV LL+K+K FKNRNVTS VAY++NML+MGFWHGLTWYYIAYGLFHGIGL+INDAW+R

Sbjct: 303 MRLVHLLIKHKTFKNNRNVTSVAYLVNMLVMGFHWGLTWYYIAYGLFHGIGLIINDAWIR 362

A related GBS gene <SEQ ID 8913> and protein <SEQ ID 8914> were also identified. Analysis of this protein sequence reveals the following:

```

bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

1233 1263 1293 1323 1353 1383 1413 1443
FVFMRLVKLLVKNKVFKNRNVTSVAYIINMLIMGFHWGLTWYYIAYGLPHGIGLVINDAWVRKKKNINKERRLAKKPLL
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

-2083-

```

FVFMRLVKVLVKNKVFKNRNVTSVAYIVNMLIMGFHWGVTWYYITYGLFHGVLVLNDAWLRRKKRLNKKERKAKNLSPL
      310      320      330      340      350      360      370

1473      1503      1533      1563      1593      1623      1653      1683
5  PENKWTYALGVFITFNVVMFSFLIFSGFLDLLWFPQPHNK**GVL*WILNQKY*QLLMTYLRWMLL*WMKTYLTQEF*T
   ||| || |||: |||||:|||||: ||| :|
   PENGWTRALGIVITFNVVMLSFLIFSGFLNDLWFADQLSKK
      390      400      410      420

```

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1850

A DNA sequence (GBSx1957) was identified in *S.agalactiae* <SEQ ID 5745> which encodes the amino acid sequence <SEQ ID 5746>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 45
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 20 bacterial cytoplasm --- Certainty=0.2611(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10143> which encodes amino acid sequence <SEQ ID 10144> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC05774 GB:AF051356 D-alanine-D-alanyl carrier protein ligase
[Streptococcus mutans]
Identities = 404/510 (79%), Positives = 465/510 (90%)

```

- 30 Query: 5 IHDMIKTIEHFAETQADFFVYDILGEVHTYQQLKVDSDSLAAHIDSLGLVEKSPVLVFGG 64
 I DMI TIE+FA+ QA+FPVY+ILGE+HTYG+LK DSDSLAAH+D L L KSPV+VFGG
 Sbjct: 6 IKDMIATIENTFAQEAEFPVYNILGEIHTYQELKADSDSLAAHLQDLDTAKSPVVVFGG 65
- 35 Query: 65 QEYEMLATFVALTKSGHAYIPVDQHSALDRIQAIMTVAQPSLIISIGFPLEVDNVPILD 124
 QEY MLA+FVALTKSGHAYIP+D HSA+RI+AI+ VA+PSL+I++ +FP++ VP++
 Sbjct: 66 QEYAMLASFVALTKSGHAYIPIDHHSALERIEAILEVAEPSLVIAVDDFPIDNLQVPVIQ 125
- 40 Query: 125 VSQVSAIFEKTPYEVTSHSVKGDDNYYIIIFTSGTGLPKGVQISHDNLLSFTNWMISDDE 184
 SQ+ IF++K Y++ H+VKGDD YYIIIFTSGTTG PKGVQISHDNLLSFTNWMI+ +
 Sbjct: 126 VSQLEEIPKQKLSYQINHAVKGDDTYIIIFTSGTGTKPKGVQISHDNLLSFTNWMINAE 185
- 45 Query: 185 FSVPERPQMLAQPPYSFDLSVMYWAPTLAMGGTLFALPKTVVNDFKLKFATINELPIQVW 244
 F+ P RPQMLAQPPYSFDLSVMYWAPTLA+GGTLFALPK + DFK+LF TIN+LPI VW
 Sbjct: 186 FATPHRPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKETADDFKQLFTTINQLPIGVW 245
- 50 Query: 245 TSTPSFADMAALLSNDNFSETLPQLTHFYFDGEELTVKTAQKLRQRFPPKARIVNAYGPTEA 304
 TSTPSF DMA+LS+DFN++ LP LTHFYFDGEELTVKTA+KLRQRFPP+ARIVNAYGPTEA
 Sbjct: 246 TSTPSFVDMAMLSDDFNAQQLPHLTHFYFDGEELTVKTAKLRQRFPPQARIVNAYGPTEA 305
- 55 Query: 305 TVALSAVAITDEMLETCKRLPIGYTKDDSPYVIDEEGHKLNGEQGEIIIAGPAVSKGY 364
 TVALSA+A+TD+MLETCKRLPIGYTK DSPT++IDE GHKL NG+QGEIII+GPAVSKGY
 Sbjct: 306 TVALSALAVTDKMLETCKRLPIGYTKPDSPFTIIDESGHKLNGQQGEIIVSGPAVSKGY 365
- Query: 365 LNNPEKTAEAFFQFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEDVSQNL 424
 LNNPE+TA AFF+FEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELE+VSQNL
 Sbjct: 366 LNNPERTAAAFFEFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEEVVSQNL 425
- 60 Query: 425 NKSQYVKSAAVAVPRYNKDHKVQNLLAYIVLKEGVRRDDFERDLDTKAIKEDLKDIMMDYM 484
 NKSQY+ SAVAVPRYNKDHKVQNLLAY+VLK+GV + FER LD+TKAIK DL+D+MMMDYM
 Sbjct: 426 NKSQYIASAAVAVPRYNKDHKVQNLLAYVVLKDGVEEQFERALDITKAIKADLQDVMMDYM 485

-2084-

Query: 485 MPSKFIYREDLPLTPNGKIDIKGLMSEVNK 514
 MPSKF+YR+DLPLTPNGKIDIKGLMSEVNK
 Sbjct: 486 MPSKFLYRKDLPLTPNGKIDIKGLMSEVNK 515

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5747> which encodes the amino acid sequence <SEQ ID 5748>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 10 INTEGRAL Likelihood = -2.28 Transmembrane 92 - 108 (91 - 108)
 INTEGRAL Likelihood = -0.85 Transmembrane 43 - 59 (41 - 59)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

20 >GP:AAC05774 GB:AF051356 D-alanine-D-alanyl carrier protein ligase
 [Streptococcus mutans]
 Identities = 365/511 (71%), Positives = 438/511 (85%)
 Query: 2 IKDMIDSIEQFAQTQADFPVYDCLGERRTYQQLKRSDSIAAFIDSLALLAKSPVLVFGA 61
 IKDMI +IE FAQ QA+FPVY+ LGE TYG+LK DSDS+AA +D L L AKSPV+VFG
 25 Sbjct: 6 IKDMIATIENTFAEQEAEPVYINILGEIHTYQELKADSDSLAAHLQDLDTAKSPVVVFG 65
 Query: 62 QTYDMLATFVALTKSGHAYIPVDVHSAAPERILAIIEIAKPSLIIAIEEFPLTIEGSLVS 121
 Q Y MLA+FVALTKSGHAYIP+D HSA ERI AI+E+A+PSL+IA+++FP+ + ++
 30 Sbjct: 66 QEYAMLASFVALTKSGHAYIPIDHHSALERIEAILEVAEPSLVIAVDDFPIDNLQVPVIQ 125
 Query: 122 LSEIESAKLAEMPYERTHSVKGDNDYIIIFTSGTTGPKGVQISHDNLLSFTNWMIEDAA 181
 S++E ++ Y+ H+VKGDD YYIIFTSGTTG+PKGVQISHDNLLSFTNWMI A
 Sbjct: 126 YSQLBEIFKQKLSYQINHAVKGDNTYIIIFTSGTTGPKGVQISHDNLLSFTNWMINAEA 185
 35 Query: 182 FDVPKQPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKELVADFKQLFTTIAQLPVGWI 241
 F P +PQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKE+ ADFKQLFTTI QLP+G+W
 Sbjct: 186 FATPHRPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKEITADFKQLFTTINQLPIGVW 245
 40 Query: 242 TSTPSFADAMLSDDFCQAKMPALTHFYFDGEELTVSTARKLFRFPSAKIINAYGPTEA 301
 TSTPSF DMAMLSDDF ++P LTHFYFDGEELTV TA+KL +RFP A+I+NAYGPTEA
 Sbjct: 246 TSTPSFVDMAMLSDDFNAQQLPHLTHFYFDGEELTVKTAKKLQRFPQARIVNAYGPTEA 305
 Query: 302 TVALSAIETREMVDNYTRLPIGYPKPSPTYIIDEKGKELSSGEQGEIIVTGAHSVSKGY 361
 TVALSA+ +T +M++ RLPIGY KPSPT+IIDE G +L++G+QGEIIV+GPAHSVSKGY
 45 Sbjct: 306 TVALSALAVTDKMLETCRRLPIGYTKPSPTFIIDESGHKLANGQQGEIIVSGPAHSVSKGY 365
 Query: 362 LNNPEKTAEAFFTFKGQPAYHTGDIGSLTEDNILLYGGRIDFQIKYAGYRIELEDVSQQL 421
 LNNPE+TA AFF F+G PAYHTGD+GS+T++ +LLYGGR+DFQIK+ GYRIELE+VSQ L
 50 Sbjct: 366 LNNPERTAAAFFEFEGLPAYHTGDLGSMIDEGLLLYGGRMDFQIKFNGYRIELEEVSNL 425
 Query: 422 NQSPMVASAVAVPRYNKEHKVQNLLAYIVVKDGVKRFDRLELTKAIKASVKDHMMSYM 481
 N+S +ASAVAVPRYNK+HKVQNLLAY+V+KDG+V+E+F+R L++TKAIK +D MM YM
 Sbjct: 426 NKSQYIASAVAVPRYNKDHKVQNLLAYVVLKDGVEEQFERALDITKAIKADLQDVMMMDYM 485
 55 Query: 482 MPSKFLYRDSLPLTPNGKIDIKTLINEVNNR 512
 MPSKFLYR LPLTPNGKIDIK L++EVN +
 Sbjct: 486 MPSKFLYRKDLPLTPNGKIDIKGLMSEVNKK 516

An alignment of the GAS and GBS proteins is shown below.

60 Identities = 374/510 (73%), Positives = 439/510 (85%)
 Query: 4 MIHDMIKTIEHFAETQADFPVYDILGEVHTYQQLKVDSDSLAAHIDSLGLVEKSPVLVFG 63
 MI DMI +IE FA+TQADFPVYD LGE TYGQLK DSDS+AA IDSL L+ KSPVLVFG
 Sbjct: 1 MIKDMIDSIEQFAQTQADFPVYDCLGERRTYQQLKRSDSIAAFIDSLALLAKSPVLVFG 60

-2085-

Query: 64 GQYEMLATFVALTKSGHAYIPVDQHSALDRIQAIMTVAQPSLIISIGEFPLEVDNVPIL 123
 Q Y+MLATFVALTKSGHAYIPVD HSA +RI AI+ +A+PSLII+I EFPL ++ + ++
 Sbjet: 61 AQTVDMLATFVALTKSGHAYIPVDVHSAPERILAIIEIAKPSLIIAIEEFPLTIEGSLV 120

Query: 124 DVSQVSAIFEEKTPYEVTSHSVKGDNDNYIIFTSGTTGLPKGVQISHDNLLSFTNWMISDD 183
 +S++ + + PYE THSVKGDNDNYIIFTSGTTG PKGVQISHDNLLSFTNWMI D
 Sbjet: 121 SLSEIESAKLAEMPYERTSHSVKGDNDNYIIFTSGTTGQPKGVQISHDNLLSFTNWMIEDA 180

Query: 184 EFSVPERPQMLAQPPYSFDLSVMYWAPTALMGGTLFALPKTVVNDFFKLIFATINELPIQV 243
 F VP++PQMLAQPPYSFDLSVMYWAPTAL+GGTLFALPK +V DFK+LF TI +LP+ +
 Sbjet: 181 AFDVPKQPQMLAQPPYSFDLSVMYWAPTALGGTLFALPKELVADFKQLFTTIAQLPVGI 240

Query: 244 WTSTPSFADMALLSNDFNSETLPQLTHFYFDGEELTVKTAQKLQRFPKARIVNAYGPTE 303
 WTSTPSFADMA+LS+DF +P LTHFYFDGEELTV TA+KL +RFP A+I+NAYGPTE
 Sbjet: 241 WTSTPSFADMALSDDFCQAKMPALTHFYFDGEELTVSTARKLFRFPKARIVNAYGPTE 300

Query: 304 ATVALSAVAITDEMLETCKRLPIGYTKDDSPYVIDEEGHKLNGEQGEIIIAGPAVSKG 363
 ATVALSA+ IT EM++ RLPIGY K DSPTY+IDE+G +L +GEQGEII+ GPAVSKG
 Sbjet: 301 ATVALSAIEITREMDNYTRLPIGYPKDPSPTYIIDEDGKELSSGEQGEIIVTGPVSKG 360

Query: 364 YLNNPEKTAEAFFQFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEDVSN 423
 YLNNPEKTAEAFF F+G PAYHTGD+GS+T++ +LLYGGR+DFQIK+ GYRIELEDVSN
 Sbjet: 361 YLNNPEKTAEAFFTFKGQPAYHTGDIGSLTEDNILLYGGRDLFQIKYAGYRIELEDVSN 420

Query: 424 LNKSQYVKSAAVAPRYNKDHQVQNLAYIVLKEGVRDDFERDLDTKAIKEDLKDIMMDY 483
 LN+S V SAVAVPRYNK+HKVQNLAYIV+K+GV++ F+R+L+LTKAIK +KD MM Y
 Sbjet: 421 LNQSPMVASAAVAPRYNKEHKVQNLAYIVVKGDKERFDRELELTKAIKASVKDHMMSY 480

Query: 484 MMPSKFIYREDLPLTPNGKIDIKGLMSEVN 513
 MMPSKF+YR+ LPLTPNGKIDIK L++EVN
 Sbjet: 481 MMPSKFLYRDSLPLTPNGKIDIKLINEVN 510

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1851

A DNA sequence (GBSx1958) was identified in *S. agalactiae* <SEQ ID 5749> which encodes the amino acid sequence <SEQ ID 5750>. This protein is predicted to be a histidine protein kinase (phoR). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.64 Transmembrane 9 - 25 (5 - 32)
 INTEGRAL Likelihood = -11.62 Transmembrane 136 - 152 (132 - 164)

----- Final Results -----
 bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB54569 GB:AJ006392 histidine kinase [Streptococcus pneumoniae]
 Identities = 105/416 (25%), Positives = 197/416 (47%), Gaps = 56/416 (13%)

Query: 7 KKFVFLTMSILIVVFLFAVSNRYNQYWDEYDAYRIVKLVAKN DY---LGIPGDEPIAL 63
 + F+F+ + + ++V+ L + NR + + ++ L+A DY L + G I
 Sbjet: 12 RDFIFILILGLFIVVTLLENNRRDNIQLQVNQKVKDLIA-GDYSKVLDMQGGSEITN 70

Query: 64 VTIDNQKMVKIQSNNTDLTNDVIEKSSSLK-----LEQGKKSRLKWSFIYSIKE----- 112
 +T + + ++ LT + +E+ S +L + G + + I I +
 Sbjet: 71 ITNNLNDLSEV----IRLTQENLEQESKRLNSILFYMTDGVLATNRRGQIIMINDTAKKQ 126

-2086-

Query: 113 ---YKDKTYTIAIMDLASYEVPYARRFLILVFT-----IFGFCLLAAVSLYLSR--- 158
 K+ +I++L E Y R LI I G L V L R
 Sbjct: 127 LGLVKEDVLNRSILELLKIEENYELRDLITQSPPELLLDSDINGEYLNLVRVRFALIRRES 186

5 Query: 159 -FIVGPVE-----TEMREKQ----FVSDASHELKTPIAAIRANVQVLEQ----QIPGNR 204
 FI G V TE +E++ FVS+ SHEL+TP+ ++++ ++ L++ +
 Sbjct: 187 GFISGLVAVLHDTTEQEKEERERRLFVSNVSHELRTPITSVKSYLEALDEGALCETVAPD 246

10 Query: 205 YLDHVVSETKRMEFLIEDLLNLSRLDEKRSKVNFKKLNSVLCQEVLLTYESLAYEEEEKC 264
 ++ + ET RM ++ DLL+LSR+D S ++ + +N + +L ++ + +E++
 Sbjct: 247 FIKVSLDETNRMMRMVTDLLHLSRIDNATSHLDVELINFTAFITFILNRFDKMKGQKEKEK 306

15 Query: 265 LNDTIED----DVWIVGEESQIKQILIIILLDNAIRHSLSKSAIQFSLKQARKKAILTISN 320
 + + D +W+ + ++ Q++ +L+NAI++S I +K + IL+IS+
 Sbjct: 307 KYELVRDYPINSIWMEIDTDKMTQVVDNIIINNAIKYSPDGGKITVRMKTTEQMIILSISD 366

20 Query: 321 PSATYSKEVMDNLFERFYQAKDDHADSLS---FGLGLSLAKAIVERHKGRIAYQE 373
 K+ + +F+RFY+ D A S + GLGLSLAK I+++HKG I A E
 Sbjct: 367 HGLGIPKQDLPRIFDRFYRV--DRARSRAQGGTGLGLSLAKEIKQHKGFIAKSE 420

A related sequence was also identified in GAS <SEQ ID 9131> which encodes the amino acid sequence <SEQ ID 9132>. Analysis of this protein sequence reveals the following:

>>> Seems to have an uncleavable N-term signal seq
 25 INTEGRAL Likelihood =-11.30 Transmembrane 9 - 25 (4 - 33)
 INTEGRAL Likelihood =-10.35 Transmembrane 161 - 177 (154 - 190)
 PERIPHERAL Likelihood = 4.35 142

----- Final Results -----
 30 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/406 (23%), Positives = 190/406 (46%), Gaps = 31/406 (7%)

35 Query: 1 MFSDLRKKFVFLTMSILIVVFLFAVSNRYNQWDEYDAYRIVKLVAKNLYLGIPGDEP 60
 MF+ +R +F+ + + +++ + + N Y + + RI+ L++ N +PG
 Sbjct: 10 MFNRIRIRFIMIASIAIFILSSIVGIINTARCYQSQQEINRILHLISSNKGK-LPGTTE 68

40 Query: 61 IAL-----VTIDNQKMKVQIS-----NNTDLTNDVIEKSSLKLE-----QGK 98
 + ++ D+ + S N L+++ S+L E + K
 Sbjct: 69 SSKRLGTLKLSDSLQFRYYSVIFNANGHLLSSNTANISALDREEAQYFARLFAKSGEK 128

45 Query: 99 KSRKWSFIYS--IKEYDKTYTIAIMDLASYEVPYARRFLILVFTIFG-FCLLAAVSLY 155
 S + + +YS I + ++ + I+D Y + V FG F +
 Sbjct: 129 GSYRHQDSVYSYLITQLPNEEKLVVILDITTFYFRSVGDLLAVSVMLAFGGFIFVVLVSL 188

50 Query: 156 LSRFIVGPVETEMTREKQFVSDASHELKTPIAAIRANVQVLEQQIPGNRYLDHVVSETKR 215
 S ++ P ++++F+++A HELKTP+A I AN +++E + + + KR
 Sbjct: 189 FSGMVIKPFVQNYEKQRRFITNAGHELKTPLAIISSANNELVELMTGESEWTKSTSDQVKR 248

55 Query: 216 MEFLIEDLLNLSRLDEKRSKVNFKKLNSVLCQEVLLTYESLAYEEEEKCLNDTIEDDVI 275
 + LI ++ L+RL+E+ V ++ S + Q+ ++SL ++ K + TI+ ++ I
 Sbjct: 249 LTGLINQMITLARLEEQPDVV-LHMVDFSALQAADAFKSLVLKDGKRFDLTIQPNIMI 307

60 Query: 276 VGEESQIKQILIIILLDNAIRHSLSKSAIQFSLK---QARRKAILTISNPSATYSKEVMDN 332
 EE + +++ IL+DNA ++ K ++ SL + R++A L +SN
 Sbjct: 308 KAEKSLFELVTILVDNANKYCDPKGLVKVSLITIGRRRKRAKLEVSNTYLEGKSIDYSR 367

Query: 333 LFERFYQAKDDH-ADSLSFGLGLSLAKAIVERHKGRIAYQEKDQL 377
 FERFY+ + H + +G+GLS+A+++V+ KG I + D +
 Sbjct: 368 FFERFYREDESHNSKEKGYGIGLSMAESMVKLFGKTTITVNYKNDI 413

-2087-

A related GBS gene <SEQ ID 8915> and protein <SEQ ID 8916> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

McG: Discrim Score: 17.50

GvH: Signal Score (-7.5): -2.9

Possible site: 26

```
>>> Seems to have an uncleavable N-term signal seq
```

```
ALOM program      count: 2 value: -13.64 threshold:  0.0
```

| | | | |
|----------|--------------------|---------------|------------------|
| INTEGRAL | Likelihood =-13.64 | Transmembrane | 9 - 25 (5 - 32) |
|----------|--------------------|---------------|------------------|

INTEGRAL Likelihood =-11.62 Transmembrane 136 - 152 (132 - 164)

| | | | |
|------------|--------------|------|-----|
| PERIPHERAL | Likelihood = | 2.49 | 345 |
|------------|--------------|------|-----|

modified ALOM score: 3.23

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.6456 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

28.3/57.2% over 371aa

Listeria monocytogenes

GP|6117973| LisK Insert characterized

ORF00341 (631 - 1452 of 1785)

GP|6117973|gb|AAF03933.1|AF139908_3|AF139908(105 - 476 of 483) LisK {Listeria monocytogenes}

```
%Match = 8.4
```

```
%Identity = 28.2   %Similarity = 57.1
```

Matches = 79 Mismatches = 113 Conservative Sub.s = 81

459 489 519 549 579 609 639 669
VKLVAKNDYLGIPGDEPIALVTIDNQMKVIQSNNTDLTNDVIEKSSSLKLLQGKKSRKWKSFYISIKKEYKDKTYTIAIM
: | : | | : | : : | : | : | | : | :
QGIGQMLLNEEPEVKELLATTSTLTNQDLTDNEEIKYLFNNDKTVNRLQDQVINLYDKDGHFINKYFYSRSQDITSI
50 60 70 80 90 100 110

[illegible]

783 813 843 873
-----VGPVETEMTREKQFVSDASHELKTPIAAIRA
: :|| :||| |||||:|: :
LLAQNFLNPLTRLARTMNDIRKNGFQRIETKTSNRDEIGELTVVFNDMMTRIETSFEEQKQFVEDASHELTRFPVQIMEG
210 220 230 240 250 260 270

918 948 978 1008 1038 1068 1098
 NVQVLEQ---QIPG--NRYLDHVVSETKRMEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEKCLN
 ::::| : : | : ::| :|| :|::| :|| : : : : :| :| : ||
 HLKLLTRWGKDDPAVLDES LNASLT ELMKKLVQEMDLDSRAEQISQPKELQITDVNATVEQVRNFE-VMYENFTFTL
 290 300 310 320 330 340 350

1128 1158 1188 1218 1248 1278 1308 1335
DTIEDDVWIVGEESEIQIKQILITLLDNAIRHLSKSAIQFSLKQARRKAILTISNPSAIYSKEVMDNLFERFYQA-KDDHA
: |: : : ::|||::|||:: : : : : : : : |:| :| :| ||| :|
KEDDTDLRALIQHNHLEQILIIIMDNVAVKSYSGDGTVEVDMHYVKEQKQIHDVRDYGEGISQEEIDKIFNRFYRVVDKARSR
370 380 390 400 410 420 430

1365 1395 1425 1452 1482 1512 1542 1572
 DLSFGLGLSIAKAIVERHKGGRIRAYQEKDQ-LRLEVQLPIDGFNTMIN*RKNDETIFIFYW*NVIILRYFIVTNLLF
 : |||:| | : | : | | | : :: | :

-2088-

EKGGNGLGLAIAKQLVEGYLGITINAVSEPDKGTTIKITLPIIEPKSK
 450 460 470 480

SEQ ID 5750 (GBS34) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 9; MW 69kDa).

5 GBS34-GST was purified as shown in Figure 193, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1852

10 A DNA sequence (GBSx1959) was identified in *S.agalactiae* <SEQ ID 5753> which encodes the amino acid sequence <SEQ ID 5754>. This protein is predicted to be two-component response regulator (regX3). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1986(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
 Identities = 98/223 (43%), Positives = 145/223 (64%), Gaps = 5/223 (2%)

25 Query: 2 RLLVVEDEKSIAEAIQALLADKGYSDLAFDGDDGLEIYLTGLYDLVLLDIMLPKRSGLS 61
 R+L++EDEK IA +Q L +GY D AF G DGLE +DLVLLD+MLP+ SGL
 Sbjct: 3 RILIIEDEKKIARVLQLELEHEGYETDAAFSGSDGLETFOAHAWDLVLLDVMLPELSGLE 62

30 Query: 62 VLKRVREAGLETPIIFLTAKSQTIDKVNGLDLGADDYITKPFADDELLARIR--LRTRQS 119
 VL+R+R TPII LTA++ DKV+GLDLGA+DYITKPF +ELLAR+R LRT Q+
 Sbjct: 63 VLRRIRMTDPVTPPIILLTARNIPDKVSGLDLGANDYITKPFIEELLARVRACLRVTQT 122

35 Query: 120 SLIRANQLRLGNIRLNTDSHELESKESSVKLSNKEFLLMEVFMNAKQIIPKNQLISKVW 179
 + L + +N + +++ +++L+ KEF L+ F++N Q++ + Q+++ VW
 Sbjct: 123 RERVEDTLMFQELTINEKTRDVRQNETIELTPKEFELLVFFIKNGQVLSREQILTINW 182

Query: 180 GPSDNSEYNQLEVFISFLRKKLRFLKADIEIITTKGFGYSLEE 222
 G + N ++V++ +LRKKL +A + T +G GY L+E
 Sbjct: 183 GFDYGDNTNVIDVYVRYLRKKLSLSEA---LQTVRGVGYRLKE 222

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1853

A DNA sequence (GBSx1960) was identified in *S.agalactiae* <SEQ ID 5755> which encodes the amino acid sequence <SEQ ID 5756>. This protein is predicted to be 50S ribosomal protein L34-related protein.

45 Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5923(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2089-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22660 GB:U32781 ribosomal protein L34 (rpL34) [Haemophilus influenzae Rd]
Identities = 32/44 (72%), Positives = 37/44 (83%)

5 Query: 1 MKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLASRRRKGRKVL 44
MKRT+QPS ++R R HGFR RM+TKNGR+VLA RR KGRK LSA
Sbjct: 1 MKRTFQPSVLKRSRTHGFRARMATKNGRQVLARRRAKGRKSL 44

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5757> which encodes the amino acid sequence <SEQ ID 5758>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.5385(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 42/44 (95%), Positives = 44/44 (99%)
Query: 1 MKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLASRRRKGRKVL 44
+KRTYQPSKIRRQRKHGFRHRMSTKNGRRVLA+RRRKGRKVL 44
Sbjct: 1 VKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLAARRRKGRKVL 44

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1854

30 A DNA sequence (GBSx1961) was identified in *S.agalactiae* <SEQ ID 5759> which encodes the amino acid sequence <SEQ ID 5760>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -5.79 Transmembrane 122 - 138 (115 - 141)
INTEGRAL Likelihood = -4.35 Transmembrane 19 - 35 (15 - 40)
35 ----- Final Results -----
bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF95990 GB:AE004350 conserved hypothetical protein [Vibrio cholerae]
Identities = 79/145 (54%), Positives = 117/145 (80%)

45 Query: 1 MKTFVNNASKTVLSLWFGVMPTIMTVGTIALIISVSTPIFKILGTPFLPFLLELLGIPEAD 60
+++ + + + FGV+P +M +GTIAL+I+ T +F +LG PF+PFLELLG+PEA
Sbjct: 314 VQSVIGEGIRNAVDMVFGVLPVVMGLGTIALVIAEYTSVFSLLGQPFIPFLLELLGVPEAT 373

50 Query: 61 IASQTMIVGFSDMVVPSIMAAEIHSEMTFRFIVATVSIVQLIYMSETGAVILGSKIPINIL 120
AS+T++VGF+DM +P+I+AA I +EMTRF++A +S+ QLIYMSE GA++LGS+IP+NI+
Sbjct: 374 AASKTIVVGFADMFIPAILAASIDNEMTRFVIAAMSVTQLIYMSEVGALLLGSRI PVNIV 433

55 Query: 121 ELFIIFIERTIISLPPIIVLMAHLFF 145
ELF+IFI RT+I+LP+I +AHL F
Sbjct: 434 ELFVIFILRTLITLTPVIAVAHLFF 458

No corresponding DNA sequence was identified in *S.pyogenes*.

-2090-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1855

A DNA sequence (GBSx1962) was identified in *S.agalactiae* <SEQ ID 5761> which encodes the amino acid sequence <SEQ ID 5762>. This protein is predicted to be D,D-carboxypeptidase (dacA-2). Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2443(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 A related GBS nucleic acid sequence <SEQ ID 9485> which encodes amino acid sequence <SEQ ID 9486> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10945> which encodes amino acid sequence <SEQ ID 10946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAA67776 GB:X99400 D,D-carboxypeptidase [Streptococcus pneumoniae]
Identities = 193/383 (50%), Positives = 282/383 (73%), Gaps = 6/383 (1%)

Query: 1 MAVDLDSGKILYEK DANKPAAIASLTKIMTVVMVYKEIDNGNLKWN TKVNISDYPYQLTR 60
+AV+ ++GKILYEKDA +P IAS+TK++TVY+VY+ ++NG++ +T V+ISDYPYQLT

25 Sbjct: 33 IAVEANTGKILYEKDATQPVEIASITKLITVYLVYEALENG SITLSTPVDISDYPYQLTT 92

Query: 61 ESDASNVPLEKRRYTVKQLVDAAMISSANSAAIALAEHISGTESKFVDKMTAQLEKWKGIH 120

S+ASN+P+E R YTV++L++A ++SSANSAAIALAE I+G+E FVD M A+L +WGI

30 Sbjct: 93 NSEASNIPMEARNYTVEEELLEATLVSSANSAAIALAEKIAGSEKDFVDMMRKLEWGIQ 152

Query: 121 DSHLVNASGLNNSMLGNHIYPKSSQNDENKMSARDIAIVAYHLVNEYP SILKITSKSVAK 180

D+ +VN +GLNN LG++IYP S +++ENK+SA D+AIVA +L+ +YP +L+IT K +

35 Sbjct: 153 DATVVTNTGLNNETLGDNIYPGSKKDEENKLSAYDVAIVARNLIKYPQVLEITKKPSST 212

Query: 181 FDKDIMHSYNYMLPDMVPVFRPGITGLKTGTTEL AGQSFIATSTESGMRLLTVMHADKAD 240

F + S NYML MP +R G GLKTGTT+ AG+SF+ T+ E GMR++TV+++AD D

40 Sbjct: 213 FAGMTITSTNYMLEGMPAYRGGFDGLKTGTTDKAGESFVGTTVKEGMRVITVVLNADHQD 272

Query: 241 KDKYARFTATNSLLNYITNTYEPNLVLAKGAAYKGKEASVRDGKEQSVI AVAKNDLKVVQ 300

+ YARFTAT+SL++YI++T+ ++ +G AY+ +A V+DGKE +VIAVA D+ +++

45 Sbjct: 273 NNPYARFTATSSLMYIISFTFLRKIVQQGDAYQDSKAPVQDGKEDTVI AVAPEDIY LIE 332

Query: 301 KKNITKQNQLKINF---KKELTAPITKKENLGKAYYVDL NKVGKGYLIKE-PSVHLVAKD 356

+ + Q+ + F K + AP+ +G Y D + +G+GY+ E PS +VA

50 Sbjct: 333 R--VGNQSSQSQVQTPDSKAIPAPLEAGTVVGHLTYEDKDLIGQGYIT TERPSFEMVADK 390

Query: 357 SIERSFFLKVWNNHFVRYVNEKL 379

IE++FFLKVWNN FVR+VNEKL

55 Sbjct: 391 KIEKAFFLKVWNNQFVRVNEKL 413

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5763> which encodes the amino acid sequence <SEQ ID 5764>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

55 ----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

-2091-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 176/380 (46%), Positives = 257/380 (67%), Gaps = 3/380 (0%)

Query: 1 MAVDLD SGKILYEKDANKPAAIASLTKIMTVVMVYKEIDNGLKWN TKVNISDYPYQLTR 60
 +AVDL+SGK+LYEKDA + +AS++K++T Y+VYKE+ G L W++ V IS+YPY+LT
 10 Sbjct: 33 IAVDLESGKVLYEKDAKEVVPVASVSKLLTTYLVYKEVSKGKLNWDSFVTISNYPYELTT 92

Query: 61 ESDASNVPLEKRRYTVKQLVDAAMISSANSAAIALAEHISGTESKFVDKMTAQLEKWGIH 120
 SNVPL+KR+YTVK+L+ A ++++ANS AIALAE I GTE KFVDKM QL +WGI
 15 Sbjct: 93 NYTISNVPLDKRKYTVKELLSALVNNNANSPALAEKIGGTEPKFVDKMKQLRQWGIS 152

Query: 121 DSHLVNASGLNNSMLGNHIYPKSSQNDENKMSARDIAIVAYHLVNEYPSILKITSKSVAK 180
 D+ +VN++GL N LG + YP + +DEN A D+AI+A HL+ E+P +LK++SKS
 20 Sbjct: 153 DAKVVNSTGLTNHFLGANTYPNTEPDDENCFCATDLAIARHLLLEFPEVLKLSKSSTI 212

Query: 181 FDKDIMHSYNYMLPDMFVFRPGITGLKTGTTEL AGQSFIATSTESGMRLLTVMHADKAD 240
 F ++SYNYML MP +R G+ GL G ++ AG SF+ATS E+ MR++TV+++AD++
 25 Sbjct: 213 FAGQTIYSYNYMLKGMPCYREGVDGLFVGYSKKAGASFVATSVENQMRVITVVLNADQSH 272

Query: 241 KDKYARFTATNSLLNYITNTYEPNLVLAKGAAYKGKEASVRDGEQSVIAVAKNDLKVVQ 300
 +D A F TN LL Y+ ++ ++ K V D E++V VA+N L ++
 30 Sbjct: 273 EDDLAI FKT TNQLLQYLLIN FQK VQLIENNKPV--KTLVVLDSPEKTVKLVAQNSLFFIK 330

Query: 301 KKNITKQNQLKINFKKE-LTAPITKKENLGKAYYVDLNVGKG YLIKEPSVHLVAKDSIE 359
 + +N + I K + AP++K + LG+A D + +G+GYL PS++L+ + +I
 35 Sbjct: 331 PIHTKTKNTVHITKKSSTMIAPLSKGQVLGRATLQDKHLIGQGYLDTPPSINLILQKNIS 390

Query: 360 RSFFLKVVWNNHFVRYVNEKL 379
 +SFFLKVWVN FVRYVN L
 40 Sbjct: 391 KSFFLKVVWNNRFVRYVNTSL 410

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1856

A DNA sequence (GBSx1963) was identified in *S.agalactiae* <SEQ ID 5765> which encodes the amino acid sequence <SEQ ID 5766>. This protein is predicted to be penicillin binding protein 4 (pdp4) (dacA-1).

40 Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-12.58 Transmembrane 368 - 384 (363 - 394)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA60582 GB:X87104 penicillin binding protein 4 [Staphylococcus aureus]
 Identities = 117/333 (35%), Positives = 188/333 (56%), Gaps = 8/333 (2%)

55 Query: 5 IVSFLCILLSLTCVNSVQAEHKKDIMQITREAGY-DVKDINKPKASIVIDNKGHILWEDN 63
 I+ LC+ LS+ + A +Q + GY + +P +++ + G +L++ N
 Sbjct: 7 IIII LCLFLSIMTPYAQAANS DVPVQAANQYGYAGLSAAYEPTSAVNVSQTGQLLYQYN 66

60 Query: 64 ADLERDPASMSKMF TLYLLFEDLAKGKTS LNTT V TATETDQAISKIYEISNNNIHAGVAY 123
 D + +PASM+K+ T+YL E + KG+ SL+ TVT T + +S + E+SN ++ G +

-2092-

Sbjct: 67 IDTKWNPASMTKMLTMYLTLEAVNKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYPGQVW 126

Query: 124 PIRELITMTAVPSSNVATIMIANHLSQNNPDAFIKRINETAKKLGMTKTHFYNPAGAVAS 183
I +L+ +T SSN A +++A +S+N D F+ +N AK +GM THF NP+GA S

5 Sbjct: 127 TIADLLQITVSNSSNAAALILAKKVSKNTSD-FVDLMNNKAKAIGMKNTHFVNPTGAENS 185

Query: 184 AFNGLYSPKEYDNNATNVTTARDLSILTYHFLKKYPDILNYTKYPEVKAMVGTPYEETFT 243
++P +Y + VTTARD +IL H +K+ P IL++T K + T + T+

10 Sbjct: 186 RLR-TFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFT-----KQLAPTTHAVTYY 239

Query: 244 TYNYSTPGAKFGLEGVDGLKTGSSPSAFAFNAFNAVTAKRQNTRLITVVLGVGDWSDQDGEYY 303
T+N+S GAK L G DGLKTGSS +A +N +T KR R+ V++G GD+ + GE

Sbjct: 240 TFNFSLEGAKMSLPCTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLGGEKQ 299

15 Query: 304 RHPFVNALVEKGFDAKNISSKTPVLKAVKPKK 336
R+ NAL+E+ F K + + + + KK

Sbjct: 300 RNMMGNALMERSFDQYKYVKILSKGEQRINGKK 332

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5767> which encodes the amino acid sequence <SEQ ID 5768>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -15.18 Transmembrane 371 - 387 (364 - 392)

----- Final Results -----

bacterial membrane --- Certainty=0.7071(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA62899 GB:X91786 penicillin-binding protein 4 [Staphylococcus aureus]

Identities = 119/328 (36%), Positives = 184/328 (55%), Gaps = 19/328 (5%)

35 Query: 6 ILTIFFICF--SVMPLVHAEDVMDIT-----RQAGYT-VSEVNRPKSSIVVDANSSDIL 57
+++I +C S+M D+T Q GY +S P S++ V + + +L

Sbjct: 4 LISIIILCLTSLIMTPYAQATNSDVTVPQAANQYGYAGLSAAEPTSANVNV-SQTGQLL 62

40 Query: 58 WQDNIDIPRDPASMSKMFILYILFEELAKGKITMDTTITATPTDQAIANIYEISNNNIVA 117
+Q NID +PASM+K+ T+Y+ E + KG++++D T+T T + ++ + E+SN +

Sbjct: 63 YQYNIDTKWNPASMTKMLTMYLTLEAVNKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYP 122

45 Query: 118 GVAYPIRDLITMTAVPSSNAATVMIANYSNNDASAFIDRVNATAKQLGMTNTHFSNAG 177
G + I DL+ +T SSNAA +++A +S N S F+D +N AK +GM NTHF N +G

Sbjct: 123 GQVWTIADLLQITVSNSSNAAALILAKKVSKN-TSDFVDMNNKAKAIGMKNTHFVNPTG 181

Query: 178 AAAQAFQGYNPNTKYDLSASNITTARDLSKLLYAFLLKKYPEIISFTNKSVVHTMVGTPYE 237
A + + PTKY +TTARD + L +K+ P+I+ FT + T+ T

50 Sbjct: 182 AENSRLR-TFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFTKQLAPTTLAVT--- 237

Query: 238 EEFHTYNHSLPDNQFGMKGVGDGLKTGSSPSAFAFNAMITAKRGKTRLITIVMGVGDWSDQN 297
++T+N SL + + G DGLKTGSS +A +N IT KRGK R+ ++MG GD+ +

55 Sbjct: 238 --YYTFNFSLEGAKMSLPCTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLG 295

Query: 298 GEFYRHPFVNALTEKGF---KDSKTLK 322
GE R+ NAL E+ F K K LSK

Sbjct: 296 GEKQRNMMGNALMERSFDQYKYVKILSK 323

An alignment of the GAS and GBS proteins is shown below.

Identities = 226/382 (59%), Positives = 289/382 (75%), Gaps = 7/382 (1%)

Query: 12 LLSLTCVNSVQAEHKDIMQITREAGYDVKDINKPKASIVID-NKGHILWEDNADLERDP 70
+ + C + + +D+M ITR+AGY V ++N+PK+SIV+D N ILW+DN D+ RDP

| | | | |
|----|------------|--|-----|
| | Sbjct: 9 | IFTFICFSVMPLVHAEDVMDITRQAGYTVSEVNRPKSSIVVDANSSDILWQDNIDIPRDP | 68 |
| | Query: 71 | ASMSKMFTLYLLFEDLAKGKTSLNTTVTATETDQAIKIYEISNNNIHAGVAYPIRELIT | 130 |
| 5 | Sbjct: 69 | ASMSKMFTLY+LFE+LAKGK +++TT+TAT TDQAI+ IYEISNNNI AGVAYPIR+LIT | 128 |
| | Query: 131 | MTAVPSSNVAITMIANHLSONNPDAFIKRINETAKKLGMTKTHFYNPSSGAVASAFNGLYS | 190 |
| | Sbjct: 129 | MTAVPSSN AT+MIAN+LS N+ AFI R+N TAK+LGMT THF N SGA A AF G Y+ | 188 |
| 10 | Query: 191 | PKEYDNNAITNVITARDLSILTYHFLKKYPDILNNTKYPEVKAMVGTPEETFTTYNSTP | 250 |
| | Sbjct: 189 | P +YD +A+N+TTARDLS L Y FLKKYP+I+++T V MVGTPYEE F TYN+S P | 248 |
| 15 | Query: 251 | GAKFGLGEGVDGLKTGSSPSA AFNALVTAKRQNTRLITVVLGVGDWSDQDGEYYRHFPVNA | 310 |
| | Sbjct: 249 | +FG++GVDGLKTGSSPSA AFNA++TAKR TRLIT+V+GVGDWSDQ+GE+YRHFPVNA | 308 |
| 20 | Query: 311 | LVEKGFKDAKNISSKT-PVLKAVKPKKEVTKTKTKSIQE--QPOTKEQWWTKTDQFIQSH | 367 |
| | Sbjct: 309 | L TEKGFKDSKTLSSKKARQKLEKLPVQ---TKKETSSQOHFKATKKQSYLVERVEDFMNHN | 365 |
| 25 | Query: 368 | FVSILIVLGTIAILCLLAGIVL | 389 |
| | Sbjct: 366 | +LI L I LL +V+ | 387 |

```

30 Lipop: Possible site: -1 Crend: 9
    McG: Discrim Score: -14.02
    GvH: Signal Score (-7.5): -2.54
        Possible site: 60
    >>> Seems to have no N-terminal signal sequence
    ALOM program count: 1 value: -12.58 threshold: 0.0
35 INTEGRAL Likelihood =-12.58 Transmembrane 339 - 355 ( 334 - 365)
    PERIPHERAL Likelihood = 1.38 99
    modified ALOM score: 3.02

40 *** Reasoning Step: 3

    ----- Final Results -----
        bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

```

ORF01254(301 - 1286 of 1698)
EGAD|40430|42591(32 - 419 of 431) penicillin binding protein 4 (pdp4) {Staphylococcus aureus} GP|1125682|emb|CAA60585.1||X87105 penicillin binding protein 4 {Staphylococcus aureus} GP|1125686|emb|CAA60582.1||X87104 penicillin binding protein 4 {Staphylococcus aureus}
%Match = 17.3
%Identity = 36.3 %Similarity = 59.6
Matches = 123 Mismatches = 130 Conservative Sub.s = 79

264      294      324      351      381      411      441      471
FPLHFIIPDLCKLCAS*RHKDIMQITREAGY-DVKDINKPKASIVIDNKGCHILWEDNADLERDPASM SKMFTLYLLFEDL
          :|   : ||   :    : ::   : |::: | | : :||| |: : || : | :
ILCLTLSIMTPYAQAANS D VTPVQAANQYG YAGLSAA YEPTSAVNVSQTGQLLYQYNIDTKWNPASMTKLM TMYLTLEAV
        20         30         40         50         60         70         80

501      531      561      591      621      651      681      711
AKGKTS LNTT VTATETDQAISKIYEISNNNIHAGVAYP IRELITMTAVPSSNVATIMIANHLSONNPDAFIKRINETAKK
||: ||: ||| } : : | : :|| : : | : :| ||| | ::| :|: | | : : ||
NGKQLSLDDTVTMTNK EYIMSTLP ELSNTKL YPGQVWTIADLLQITVSNSSNAALILAKKVSKNTSD-FVDLMNNKAKA

```

-2094-

```

      100      110      120      130      140      150      160
741      771      801      831      861      891      921      951
5  LGMTKTHFYNPSPGAVASAFNGLYSPKEYDNNATNVTTARDLSILTYHFLKKYPDILNNTKYPEVKAMVGTPTYEETFTTYN
   :|| ||| ||:| | : : : ||||| :| | :| : ||::| : | : | :|
   IGMKNTHFVNPTGAENSR-LRTFAPTKEYKDQERTVTTARDYAILDLHVIKETPKILDFTK-----QLAPTTHAVTYTYTFN
      180      190      200      210      220      230      240

981      1011      1041      1071      1101      1131      1161
10 YSTPGAFLGLEGVDGLKTGSSPSAFAFNALVTAKRQNTRLITVVLGVGDWSDQDGEYYRHPFVNALVEKGFKDAK-----
   :| ||| | | ||||| | :| :| :| || :| :| ||: : || | : |||:| : |
   FSLEGAKMSLPGTDLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLGGEKQRNMGMNLMERSFDQYKYVKILS
      260      270      280      290      300      310      320

1179      1209      1239      1266
15 -----NISSKTPVLKAVKPKKEVTKTKTKSI-QEQPQ
   | : | : : | : | ||:| :|
KGEQRINGKKYYVENDLYDVLPSDFSKKDYKLVEDGKVHADYPREFINKDYRPPTVEVHQPIIQKANTVAKSMWEEHP-
      340      350      360      370      380      390      400

1296      1326      1356      1386      1416      1446      1476      1506
20 TKEQWWTKTQDFIQSHFVSILIVLGTIAILCLLAGIVLLIKRSR**LC*YKSPLHQ*HRGFLLSLEIFN*PTEPSIS*EI
   : : | ||:| : :
-----LFTIIGGACLVAGLALIVHMIINRLFRKRK
      410      420      430
25

```

SEQ ID 8918 (GBS379) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 5; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 3; MW 68.9kDa).

GBS379-GST was purified as shown in Figure 212, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1857

A DNA sequence (GBSx1964) was identified in *S.agalactiae* <SEQ ID 5769> which encodes the amino acid sequence <SEQ ID 5770>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 49
   >>> Seems to have no N-terminal signal sequence

----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.4039(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45 >GP:CAB15256 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
   Identities = 316/459 (68%), Positives = 386/459 (83%)

Query: 14 DLGEYKFGFHDDVKPIYSTGKGLNEAVIRELSAAKGEPEWMLDFRLKSLETFNKMPMQTW 73
      D+GEYK+GFHD I+ + +GL + ++ E+S K EP+WMLDFRLKSLE F MPM W
Sbjct: 7 DIGEYKYGFHDKDVSIFRSERGLTKEIVEEISRMKEEPQWMLDFRLKSLEHFYNMMPQW 66

50 Query: 74 GADLSIDIDFDDIIYYQKASDKPARDWDDVPEKIKETFERIGIPEAERAYLAGASAQYESE 133
      G DL+ ++FD+I YY K S++ R WD+VPE+IK+TF+++GIPEAE+ YLAG SAQYESE
Sbjct: 67 GGDLSNLSNFDEITYYVKPSERSERSWDEVPKEIKQTFDKLGIPEAQKYLAVGSAQYESE 126

55 Query: 134 VVYHNMKEEYDKLGIVFTDTSALKEYPELFKKYFAKLVPPDTNKLALNSAVWSSGGTFI 193
      VVYHNMKE+ + GIVF DTDSALKE ++F++++AK++PPTDNK AALNSAVWSSG+FI
Sbjct: 127 VVYHNMKEDLEAQGIVFKDTSALKENEDIFREHWAKVIPPTDNKFAALNSAVWSSGGSFI 186

```

-2095-

Query: 194 YVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTAPTYSSNSLHAAIV 253
 YVPKGVKV+ PLQ YFRIN+EN GQFERTLIIVDE ASVHYVEGCTAP Y++NSLH+A+V
 Sbjct: 187 YVPKGVKVVETPLQAYFRINSENMGQFERTLIIVDEEASVHYVEGCTAPVYTTNSLHSAVV 246

5 Query: 254 EIFALDGAYMRYTTIQNWSNVNLTNRATAKKDATVEWIDGNLGAKT'TMKYPSVYLDG 313
 EI G Y RYTTIQNW++NVNLTNR +++AT+EWIDGN+G+K TMKYP+ L G
 Sbjct: 247 EIIIVKGGYCRYTTIQNWANNVYNLTNRATVCEENATMEWIDGNIGSKLTMKYPACILKG 306

10 Query: 314 EGARGTMLSIAFANKGQHODTGAKMIHNAPHTSSSIVSKSIAGGGKVDYRGQVTFNKDS 373
 EGARG LSIA A KQHOD GAKMIH AP+TSS+IVSKSI+K GGV YRG V F + +
 Sbjct: 307 EGARGMTLSIALAGKQHODAGAKMIHLAPNTSSTIVSKSISKQGGKVYRGIVHFGRKA 366

15 Query: 374 KKSVSHEICTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEEQLYYLMRGLSEA 433
 + + S+IECDT++MD+ S SDTIP+NEI N ++LEHEAKVSK+SEQL+YLMRGL+SE
 Sbjct: 367 EGARSNIECDTILMDDISKSDTIPYNEILNDNISLEHEAKVSKVSEQLFYLMRGLISEE 426

20 Query: 434 EATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
 EATEMIVMGF+EPFTKELPMEYAVE+NRLI +EMEGS+G
 Sbjct: 427 EATEMIVMGFIEPFTKELPMEYAVEMNRLIKFEMEGSIG 465

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5771> which encodes the amino acid sequence <SEQ ID 5772>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3780(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 445/472 (94%), Positives = 461/472 (97%)

35 Query: 1 MSEINEKVEPQPIDLGEYKFGFDDVKPIYSTGKGLNEAVIRELSAAKGEPEWMLDFRLK 60
 MS+INEKVEP+PIDLG+Y+FGFDDV+PIYSTGKGL+EAV+RELSAAK EPEWML+FRK
 Sbjct: 1 MSDINEKVEPKPIDLDGYQFGFDDVEPIYSTGKGLSEAVVRELSAAKNEPEWMLDFRLK 60

40 Query: 61 SLETFNKMPMQTWGADLSDIDFDDIIYYQKASDKPARWDDVPEKIKETFERIGIPEAER 120
 SLETFNKMPMQTWGADLSDI+FDIIYYQKASDKPAR WDDVPEKIKET+RIGIPEAER
 Sbjct: 61 SLETFNKMPMQTWGADLSDINFDDIIYYQKASDKPARSWDDVPEKIKETFDRIPIPEAER 120

45 Query: 121 AYLAGASAQYSEVVYHNMKEEYDKLGIVFTDTSALKEYPELFKKYFAKLVPPTDNKLA 180
 AYLAGASAQYSEVVYHNMK E++KLGI+FTDTSALKEYP+LFK+YFAKLVPPTDNKLA
 Sbjct: 121 AYLAGASAQYSEVVYHNMKGEFEKGLIIFTDTSALKEYPDLFKQYFAKLVPPTDNKLA 180

50 Query: 181 ALNSAVWSGGTFIYVPGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA 240
 ALNSA WSGGTFIYVPGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA
 Sbjct: 181 ALNSAAWSGGTFIYVPGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA 240

55 Query: 241 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSNVNLTNRATAKKDATVEWIDGNLGA 300
 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSNVNLTNR A DATVEWIDGNLGA
 Sbjct: 241 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSNVNLTNRARALTDATVEWIDGNLGA 300

60 Query: 301 KT'TMKYPSVYLDGEGARGTMLSIAFANKGQHODTGAKMIHNAPHTSSSIVSKSIAGGGK 360
 KT'TMKYPSVYLDG GARGTMLSIAFAN GQHODTGAKMIHNAPHTSSSIVSKSIAG GKK
 Sbjct: 301 KT'TMKYPSVYLDGPGARGTMLSIAFANAGQHODTGAKMIHNAPHTSSSIVSKSIAGSGK 360

65 Query: 361 VDYRGQVTFNKDSKKSVSHEICTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEE 420
 VDYRGQVTFNK SKKSVSHEICTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEE
 Sbjct: 361 VDYRGQVTFNKQSKKSVSHEICTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEE 420

Query: 421 QLYYLMRGLSEAEATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
 QLYYLMRGLSE+ATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG
 Sbjct: 421 QLYYLMRGLSESEATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472

-2096-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1858

A DNA sequence (GBSx1965) was identified in *S.agalactiae* <SEQ ID 5773> which encodes the amino acid sequence <SEQ ID 5774>. This protein is predicted to be nitrogen fixation protein (nifU). Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1078(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15257 GB:Z99120 similar to NifU protein homolog [Bacillus subtilis]
Identities = 72/139 (51%), Positives = 92/139 (65%)

Query: 4 SKLDNLYMAVVADHSHKPHHHGFLEGVEQVQLNNPTCGDVISLSVKFDGNIISDIAFAGN 63
+ LD LY V+ DH K+P + G L V +NNPTCGD I L++K DG+I+ D F G
Sbjct: 5 ANLDTLYRQVIMDHYNPRNKGVLNDSIVVDMNNPTCGDRIRLTMKLDGDIVDAKFEFE 64

Query: 64 GCTISTASSSMMTDAVIGKTKKEALQLADVFSKMVGQDQNPKEKLGDAEFLAGVSKFPQ 123
GC+IS AS+SMMT A+ GK E AL ++ +FS M+QG + LGD E L GVSKEP
Sbjct: 65 GCSISMASASMMTQAIGKDIETALSMKIFSDMMQKEYDDSIDLGDIKALQGVSKFPA 124

Query: 124 RIKCATLSWNALRKAIERD 142
RIKCATLSW AL K + ++
Sbjct: 125 RIKCATLSWKALEKGVAK 143

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5775> which encodes the amino acid sequence <SEQ ID 5776>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1202(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 114/146 (78%), Positives = 133/146 (91%)

Query: 1 MALSKLDNLYMAVVADHSHKPHHHGFLEGVEQVQLNNPTCGDVISLSVKFDGNIISDIAF 60
MALSKL++LYMAVVADHSHK PHHHG L+GVE VQLNNPTCGDVISL+VKFD + I DIAF
Sbjct: 1 MALSKLNHLYMAVVADHSHKPHHHGQLDGVAVQLNNPTCGDVISLTVKFDEDKIEDIAF 60

Query: 61 AGNGCTISTASSSMMTDAVIGKTKKEALQLADVFSKMVGQDQNPKEKLGDAEFLAGVSK 120
AGNGCTISTASSSMMTDAVIGK+KEEAL LAD+FS+MVQG +NP Q++LG+AE LAGV+K
Sbjct: 61 AGNGCTISTASSSMMTDAVIGKSKKEALALADIFSEMVGQENPAQKELGEAELLAGVAK 120

Query: 121 FPQRIKCATLSWNALRKAIERDNOAE 146
FPQRIKC+TL+WNAL++AI+R A+
Sbjct: 121 FPQRIKCSTLAWNALKEAIKRSANAQ 146

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2097-

Example 1859

A DNA sequence (GBSx1966) was identified in *S.agalactiae* <SEQ ID 5777> which encodes the amino acid sequence <SEQ ID 5778>. This protein is predicted to be nitrogen fixation protein (nifS) (b1680). Analysis of this protein sequence reveals the following:

```

5   Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.2453(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB15258 GB:Z99120 similar to NifS protein homolog [Bacillus subtilis]
    Identities = 240/400 (60%), Positives = 306/400 (76%), Gaps = 5/400 (1%)

    Query: 9   LKQDFPILNQLVNDEPLIYLDNAATTQKPNQVLEALRDYYQNDNANVHRGVHTLAERATA 68
            +++ FPIL+Q VN   L+YLD+AAT+QKP V+E L YY   N+NVHRGVHTL RAT
    Sbjct: 6   IREQFPILHQVNGHDLVYLDASAATSQKPRAVIETLDKYYNQYNSNVHRGVHTLGTRATD 65

20  Query: 69  QYENAREKARQFLNAKLSKEILFTRGTTTGLNWA-KFAESILERGDEVLIISIMEHHSNI 127
            YE AREK R+F+NAK   EI+FT+GTIT LN VA  +A + L+ GDEV+I+ MEHH+NI
    Sbjct: 66  GYEGAREKVRKFINAKSMAEIIIFTKGTTTSLNMVALSYARANLKPGEVVITYMEHHANI 125

25  Query: 128 IPWQQACERTGAKLVYAYLK-DGSLDLEDYFNKLSSKTKFVSLAHISNVLCVTPVKAIA 186
            IPWQQA + TGA L Y  L+ DG++ LED   ++S TK V+++H+SNVLG V P+K +A
    Sbjct: 126 IPWQQAVKATGATLKYIPLQEDGTISLEDVRET VTSNTKIVAVSHVSNVLGTVPNIKEMA 185

30  Query: 187 ERVHQVGAYMVVDGAQSAPHMAIDVQDLDCDFFALSGHKMLGPTGIGVLYGKESILDKMP 246
            + H GA +VVDGAQS PHM IDVQDLDCDFFALS HKM GPTG+GVLYGK+++L+ M
    Sbjct: 186 KIAHDNGAVIVVDGAQSTPHMKIDVQDLDCDFFALSSHKMCGPTGVGVLYGKKALLENME 245

    Query: 247 PVEFGGEMIDFVYEQSATWKELPWKFEAGTPNIAGAIAFGEALDYLTVDGMDEIHQYEQS 306
            P EFGGEMIDFV   +TWKELPWKFEAGTP IAGAI  G A+D+L ++G+DEI ++E
35  Sbjct: 246 PAEFGGEMIDFVGLYESTWKELPWKFEAGTPIIAGAIGLGAIDFLEEIGLDEISRHEHK 305

    Query: 307 LVSYVLPKLQAIDGLTIYGPSDAESHVGVIAFNLEGLHPHDVATAMDYEGVAVRAGHHCA 366
            L +Y L + + +DG+T+YGP   E  G++ FNL+ +HPHDVAT +D EG+AVRAGHHCA
    Sbjct: 306 LAAYALERFRQLDGVTVYGP---EERAGLVTFNLDVHHPHDVATVLDABGIAVRAGHHCA 362

40  Query: 367 QPLINHLGIHSAVRASFYFYNTKEDCDKLVDAIQKTKEFF 406
            QPL+  L + +  RASFY YNT+E+ DKLV+A+QKTKE+F
    Sbjct: 363 QPLMKWLDVTATARASFYLYNTEEEIDKLVEALQKTKEYF 402

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5779> which encodes the amino acid sequence <SEQ ID 5780>. Analysis of this protein sequence reveals the following:

```

    Possible site: 41
    >>> Seems to have no N-terminal signal sequence

50  ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

55 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 293/408 (71%), Positives = 349/408 (84%)

    Query: 3   LLD SYKLKQDFPILNQLVNDEPLIYLDNAATTQKPNQVLEALRDYYQNDNANVHRGVHTL 62
            LLD+  +KQDF ILNQ VNDEPL+YLDNAATTQKP VLEAL+ YYQ DNANVHRGVHTL
60  Sbjct: 1   LLD AKDIKQDFQILNQVNDEPLVYLDNAATTQKPALVLEALQSYQEDNANVHRGVHTL 60

```

-2098-

Query: 63 AERATAQYENAREKARQFLNAKLSKEILFTRGTTTGLNWVAKFAESILERGDEVLSIME 122
 AERAT +YE +R++ F++AK SKE+LFTRGTTT INWVA+FAE +L DEVLISIME
 Sbjct: 61 AERATLKYEASRQQVADFIHAKSSKEVLFTRGTTTSLNWVARFAEQVLTPEDEVLSIME 120

5 Query: 123 HHSNIIPWQQACERTGAKLVYAYLKDGLDLEDFYNKLSSKTKFVSLAHISNVLGCVTPV 182
 HH+NIIPWQQAC++TGA+LVY YLKDGLD++D NKL++KT+FVSL H+SNVLGC+ P+
 Sbjct: 121 HHANIIPWQQACQKTGARLVYVYLKDGLDMDLANKLTTKTRFVSLVHVS NVLGCINPI 180

10 Query: 183 KAIAERHVQVGAYMVVDGAQSAPHMAIDVQDLDCDFFALSGHKMLGPTGIGVLYGKESIL 242
 K IA+ H GAY+VVDGAQS PH+AIDVQDLDCDFFA S HKMLGPTG+GVLYGKE +L
 Sbjct: 181 KEIAKLAHAKGAYLVVDGAQSVPHLAIDVQDLDCDFFAFSAHKMLGPTGLGVLYGKEELL 240

15 Query: 243 DKMPPEVEFGGEMIDFVYEQSATWKELPWKFEAGTPNIAGAI AFGEALDYLTVDGMDI HQ 302
 +++ P+EFGGEMIDFVYEQ ATWKELPWKFEAGTP+IAGAI A+ YL +GM +IH
 Sbjct: 241 NQVEPLEFGGEMIDFVYEQEA TWKELPWKFEAGTPHIAGAILSAISYLRQLGMADIHA 300

20 Query: 303 YEQSLVSYVLPKLQAIDGLTIYGPSDAESHVGVIAFNLEGLHPHDVATAMDYEGVAVRAG 362
 +E L++YVLPKL+AI+GLTIYGPS + G+I+FNL+ LHPHD+ATA+DYEGVAVRAG
 Sbjct: 301 HEAELIAYVLPKLEAIEGLTIYGPSQPSARSGLISFNLDDLHPHDLATALDYEGVAVRAG 360

Query: 363 HHCAQPLINHLGIHSAVRASFYFYNTKEDCDKLVDIAIQTKKEFFNGTL 410
 HHCAQPL+++LG+ + VRASFY YNTK DCD+LV+AI K KEFFNGTL
 Sbjct: 361 HHCAQPLLSYLGVPATVRASFYIYNTKADCDRLVEAILKAKEFFNGTL 408

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1860

A DNA sequence (GBSx1967) was identified in *S.agalactiae* <SEQ ID 5781> which encodes the amino acid sequence <SEQ ID 5782>. Analysis of this protein sequence reveals the following:

30 Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.1441(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB07189 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 171/430 (39%), Positives = 267/430 (61%), Gaps = 15/430 (3%)

Query: 1 MSKEAILNFLQAKGEPTWLQELRLKAFEKIEELELPVIERVKFHRWNLG--DGILENDY 58
 + KE + +F A+ EP W +++RLK FE +E LELP ++ K WN D + E
 Sbjct: 9 IDKEYVQSFS DARNEPQWFKDIRLKG FELVETLELPKPKDKTKITSWNFTNFDHKLPEVSP 68

45 Query: 59 TANVPDFTE-----LGNNPKLVQIGTQTVLEQVPMELIEKGVVFTDFYSALEEIP 109
 A++ + + LVQ V ++ L KGV+FTD +A++E +
 Sbjct: 69 VASIDELRDEVKGLIGEASDTQNLVQRDATVVYSKLDEALKAKGVIFTDLLTAVKEHGD 128

50 Query: 110 VIERYFGK-ARPFEDRLAAYHTAYFNSGAVLYIPDNVEITQPIEGLFYQDSQS KVPFNK 168
 ++E+Y+ K A +E+RL A H A N G +Y+P NVEI P++ +F+ D++ FN
 Sbjct: 129 LVEKYMKDAVKVDENRLTALHAALVNGGTFIYVPRNVEIEVPLQSVFWFDTEKAGLFN- 187

55 Query: 169 HILLIVGKNKVS YLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRLGENVTTFI 228
 H++++ N+ ++Y+E + S G +E ANI VEV A A +++ F ++D L VIT++
 Sbjct: 188 HVIIIVAE DNSSITYVENYASFG--SEEAVANIVVEVFAGANAKVSFGAVDNLAAGVTITYV 245

60 Query: 229 SRRGRHSSDATIDWALGVMNEGNVADFDSDLIGDGS HANLKVVAASSGRQVQIDTRVT 288
 RR D++WALG MN+GN V++ + L+GD S A+ K V+ G Q Q T++
 Sbjct: 246 VRRAHVGRDSRVEWALGQMNDGNTVSENTTHLLGDNSWADTKTVSVGRGEQKQNFITQIF 305

Query: 289 NYGCNSVGHILQHGVILERGTLTFNGIGHI IKGAKGADAQQESRVLM LSDKARSDANPIL 348

-2099-

++G +S G+IL+HGV+ E T FNGI I GA + +Q RVLMLS+KAR DANPIL
 Sbjct: 306 HHGKHSEGYILKHGVMREAATSIFNGISKIEHGATKSHGEQTERVLMMLSEKARGDANPIL 365
 Query: 349 LIDENDVTAGHAASIGQVDPEDLYYLMRGLNOKTAEQLVIRGFLGTVIAEIPVKEVRDE 408
 5 LIDE+DVTAGHAAS+G++DP ++YLMSRG+++ AE+LVI GFL V+ ++P++ V++
 Sbjct: 366 LIDEDDVTAGHAASVGKIDPIQMFYLMRSGISRAEERLVIHGFLAPVVGQLPIESVKER 425
 Query: 409 MIAVIDTKLE 418
 ++ I+ K++
 10 Sbjct: 426 LVEAIERKVK 435

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5783> which encodes the amino acid sequence <SEQ ID 5784>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 387 - 403 (387 - 403)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:CAB15259 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 177/428 (41%), Positives = 267/428 (62%), Gaps = 15/428 (3%)
 Query: 3 KEKLVAFSQAHAEPALQERRLAALAEIPNLELPTIERVKFHRWNLDGDT--LTENESLA 60
 +E L +FS+ H EPAWL+ RL ALE +L +P ++ K WN + +NE L+
 30 Sbjct: 11 QEYLKSFSEKHQEPALWLNRLQALEQAEIDLMPKPKDKTKITNWNFTNFAKHTVDNEPLS 70
 Query: 61 SVPDF-----IAIGDNPKLQVGTQTVLEQLPMA--LIDKGVVFSDFYTALEEIPEVI 111
 S+ D I I + K + V L ++ L DKG V+F+D TA E +++
 35 Sbjct: 71 SLEDLTDEVKALIDIEDNKTLVQORDQTPAHLSSLSQELKDKGVIFDILTAAAREHSDLV 130
 Query: 112 EAHFGQ-ALAFDEDKLAAYHTAYFNSAAVLYVPDHLEITTPIEAIFLQSDSDVPFNKHV 170
 E +F + + DE KL A H A N A LYVP ++++ TP++A+++ +S+ FN HV
 Sbjct: 131 EKYFMKDGKVKVDEHKLTAALVNGGAFLYVPKNVQVETPVQAVYVHESNDTALFN-HV 189
 40 Query: 171 LVIAGKESKFTYLERFESIGNATQKISANISVEVIAQAGSQIKFSAIDRLGPSVTTYISR 230
 L++A S TY+E + S N + NI EVI + + + A+D L VTTY++R
 Sbjct: 190 LIVAEDHSSVTYVENYISTVNPDAVF-NIIESEVITGDNASVTYGAVDNLSSGVTTYVNR 248
 45 Query: 231 RGRLE-KDANIDWALAVMNEGNVIADFDSDLIQGSQADLKVVAASSGRQVQGIDTRVTN 289
 RG +D+ I+WAL +MN+G+ I++ ++L G G+ D K V G Q + T++ +
 Sbjct: 249 RGAARGRDSKIEWALGLMNDGDTISENTNLYGDTYGDTKTVVVGREGQTFENFTTQIIH 308
 Query: 290 YGQRTVGHILQHGVLIRGTLTFNGIGHILKDAKGADAQQESRVLMMLSDQARADANPILL 349
 +G+ + G+IL+HGV+ + + FNGIG I A A+A+QESRVLMML+AR DANPILL
 50 Sbjct: 309 FGKASEGYILKHGVMKDSASSIFNGIGKIEHGASKANAEQESRVLMMLSEKARGDANPILL 368
 Query: 350 IDENEVTAGHAASIGQVDPEDMYLMSRGLDQETAERLVIRGFLGAVIAEIPISVRQEI 409
 IDE++VTAGHAAS+G+VDP +YYLMSRG+ +E AERLVI GFL V+ E+PI V++++
 55 Sbjct: 369 IDEDDVTAGHAASVGRVDPIQLYYLMSRGIPKEEAERLVITYGFLAPVVNELPIEGVKKQL 428
 Query: 410 IKVLDEKL 417
 + V++ K+
 Sbjct: 429 VSVIERKV 436

An alignment of the GAS and GBS proteins is shown below.

Identities = 322/420 (76%), Positives = 368/420 (86%)

Query: 1 MSKEAILNFLQAKGEPTWLQELRLKAFKIEEELPVIERVKFHRWNLDGDTILENDYTA 60
 M+KE ++ F QA EP WLQE RL A E I LELP IERVKFHRWNLDGDT+ EN+ A

-2100-

Sbjct: 1 MTKEKLVAFSQAHAEPALQERRLAALAEIAPNLELPTIERVKFHRWNLGDGTLTENESLA 60

Query: 61 NVPDFTELGNPKLVQIGTQTIVLEQVPMELIEKGVVFTDFYSALEEIPEVIERYFGKARP 120
+VPDF +G+NPKLVQ+GTQTIVLEQ+PM LI+KGVVF+DFY+ALEEIPEVIE +FG+A

5 Sbjct: 61 SVPDFIAIGDNPKLVQVGTQTIVLEQLPMALIDKGVVFSDFYTALEEIPEVIEAHFGQALA 120

Query: 121 FEEDRLAAYHTAYFNSGAVLYIPDNVEITQPIEGLFYQDSQSKVPFNKHILLIVGKNAKV 180
F+ED+LAAYHTAYFNS AVLY+PD++EIT PIE +F QDS S VPFNKH+L+I GK +K

10 Sbjct: 121 FDEDKLAAYHTAYFNSAAVLYVPDHLBITTPIEAIFLQSDSDVPFNKHVLVIAGKESKF 180

Query: 181 SYLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRLGENVTTTISRGRHSSDATI 240
+YLERFESIG+ T++ SANISVEVIAQAGSQIKF++IDRLG +VTT+ISRGR DA I

Sbjct: 181 TYLERFESIGNATQKISANISVEVIAQAGSQIKFSAIDRLGPSVTTYISRGRLEKDANI 240

15 Query: 241 DWALGVMNEGNVADFDSDLIGDGSANLKVVAASSGRQVQIDTRVTNYGCNSVGHILQ 300
DWAL VMNEGNV+ADFDSDLIG GS A+LKVVAASSGRQVQIDTRVTNYG +VGHILQ

Sbjct: 241 DWALAVMNEGNVIADFDSDLIGQGSQADLKVVAASSGRQVQIDTRVTNYGQRTVGHILQ 300

20 Query: 301 HGVILERTLTTFNGIGHIIGKAGGADAQQESRVLMLSDKARSDANPILLIDENDVTAGHA 360
HGVILERTLTTFNGIGHI+K AKGADAQQESRVLMLSD+AR+DANPILLIDEN+VTAGHA

Sbjct: 301 HGVILERTLTTFNGIGHILKDAKGADAQQESRVLMLSDQARADANPILLIDENEVTAGHA 360

Query: 361 ASIGQVDPEDLYLMSRGLNQKTAEQLVIRGFLGTVIAEIPVKEVRDEMIVIDTKLEKR 420
ASIGQVDPED+YYLMSRGL+Q+TAE+LVIRGFLG VIAEIP+ VR E+I V+D KL R

25 Sbjct: 361 ASIGQVDPEDMYLMSRGLDQETAERLVIRGFLGAVIAEIPPSVRQEIIKVLDEKLLNR 420

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1861

30 A DNA sequence (GBSx1968) was identified in *Sagalactiae* <SEQ ID 5785> which encodes the amino acid sequence <SEQ ID 5786>. This protein is predicted to be ABC transporter, ATP-binding protein, Ycf16 family. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2253(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15260 GB:Z99120 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 180/250 (72%), Positives = 212/250 (84%)

Query: 2 SVLEIKNLHVSIEDKEILKGLNLTLKTGEIAAIMGPNGTGKSTLSAAIMGNPNYEVTAGE 61

S L IK+LHV IE KEILKG+NL +K GE A+MGPNGTGKSTLSAAIMG+P YEVT G

Sbjct: 4 STLTIKDLHVEIEGKEILKGVNLEIKGGEFHAVMGPNGTGKSTLSAAIMGHPKYEVTKGS 63

Query: 62 ILFDGEDILELEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGKADDDKISIRQFI 121

I DG+D+LE+EVDERA+ GLFLAMQYPSE+ G+TNA+F+R+A+NA + + D+IS+ +FI

Sbjct: 64 ITLDGKDVLEMEVDERAQAGLFLAMQYPSEISGVTNADFLRSAINARREEGDEISLMKFI 123

Query: 122 TKLDEKMELLMKMEEMAERYLNEGFSGGEKKRNEILQLMLLEPKFALLDEIDSGLDIDAL 181

K+DE ME L M EMA+RYLNEGFSGGEKKRNEILQL+M+EPK A+LDEIDSGLDIDAL

Sbjct: 124 RKMDENMEFLEMDPEMAQRYLNEGFSGGEKKRNEILQLMMIEPKIAILDEIDSGLDIDAL 183

Query: 182 KVVSKGVNEMRGEFGGAMIITHYQRLNLYITPDKVHVMMDGKVVLSGGPELAVRLEKEGY 241

KVVSKG+N+MR E FG ++ITHYQRLNLYITPD VHVMM G+VV SGG ELA RLE EGY

Sbjct: 184 KVVSKGINKMRSENFGLMITHYQRLNLYITPDVVHVMMQGRVVKSGGAELAQRLEAEGY 243

-2101-

Query: 242 AQIAEELGLE 251
 I +ELG+E
 Sbjct: 244 DWIKQELGIE 253

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5787> which encodes the amino acid sequence <SEQ ID 5788>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2417(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 225/255 (88%), Positives = 241/255 (94%)

Query: 1 MSVLEIKNLHVSIEDKEILKGLNLTLKTGEIAAIMGPNGTGKSTLSAAIMGNPNYEVTAG 60
 MS+LEI NLHVSIE KEILKG+NLTLKTGE+AAIMGPNGTGKSTLSAAIMGNPNYEVT G
 20 Sbjct: 1 MSILEINNHLHVSIEGKEILKGVNLTLKTGEVAAIMGPNGTGKSTLSAAIMGNPNYEVTQG 60
 Query: 61 EILFDGEDILEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGKADDDKISIRQF 120
 +IL DG +IL+LEVDERARLGLFLAMQYPSE+PGITNAEF+RAAMNAGKAD+DKIS+R F
 25 Sbjct: 61 QILLDGVNILDLEVDERARLGLFLAMQYPSEIPGITNAEFMRAAMNAGKADEDKISVRDF 120
 Query: 121 ITKLDEKMELLMKEEMAERYLNFGSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA 180
 ITKLDEKM LLGMKEEMAERYLNFGSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA
 Sbjct: 121 ITKLDEKMALLGMKEEMAERYLNFGSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA 180
 30 Query: 181 LKVVSKGVNEMRGEGFGAMIITHYQRLNLYITPDKVHVMDGKVVLSGGPELAVRLEKEG 240
 LKVVSKGVNEMRG+ FGAMIITHYQRLNLYITPD VHVMDG++VLSG LA RLEKEG
 Sbjct: 181 LKVVSKGVNEMRGKDFGAMIITHYQRLNLYITPDLVHVMDGRIVLSGDAALATRLLEKEG 240
 Query: 241 YAQIAEELGLEEYKEE 255
 35 YA IA++LG+EYKEE
 Sbjct: 241 YAGIAQDLGIEYKEE 255

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1862

A DNA sequence (GBSx1969) was identified in *S.agalactiae* <SEQ ID 5789> which encodes the amino acid sequence <SEQ ID 5790>. This protein is predicted to be RgpG (rfe). Analysis of this protein sequence reveals the following:

Possible site: 40
 45 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.10 Transmembrane 312 - 328 (308 - 336)
 INTEGRAL Likelihood = -10.03 Transmembrane 15 - 31 (6 - 41)
 INTEGRAL Likelihood = -9.82 Transmembrane 205 - 221 (197 - 226)
 INTEGRAL Likelihood = -8.60 Transmembrane 335 - 351 (329 - 358)
 50 INTEGRAL Likelihood = -7.48 Transmembrane 257 - 273 (255 - 281)
 INTEGRAL Likelihood = -5.52 Transmembrane 60 - 76 (56 - 79)
 INTEGRAL Likelihood = -5.31 Transmembrane 151 - 167 (148 - 171)
 INTEGRAL Likelihood = -4.88 Transmembrane 91 - 107 (90 - 108)
 INTEGRAL Likelihood = -4.78 Transmembrane 184 - 200 (177 - 203)
 55 INTEGRAL Likelihood = -3.13 Transmembrane 119 - 135 (119 - 135)
 INTEGRAL Likelihood = -2.97 Transmembrane 229 - 245 (229 - 250)

----- Final Results -----
 bacterial membrane --- Certainty=0.5840(Affirmative) < succ>

-2102-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8919> which encodes amino acid sequence <SEQ ID 8920> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 5.18
 GvH: Signal Score (-7.5): -6.19
 Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 9 value: -12.10 threshold: 0.0

| | | | |
|------------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -12.10 | Transmembrane | 239 - 255 (235 - 263) |
| INTEGRAL | Likelihood = -9.82 | Transmembrane | 132 - 148 (124 - 153) |
| INTEGRAL | Likelihood = -8.60 | Transmembrane | 262 - 278 (256 - 285) |
| INTEGRAL | Likelihood = -7.48 | Transmembrane | 184 - 200 (182 - 208) |
| INTEGRAL | Likelihood = -5.31 | Transmembrane | 78 - 94 (75 - 98) |
| INTEGRAL | Likelihood = -4.88 | Transmembrane | 18 - 34 (17 - 35) |
| INTEGRAL | Likelihood = -4.78 | Transmembrane | 111 - 127 (104 - 130) |
| INTEGRAL | Likelihood = -3.13 | Transmembrane | 46 - 62 (46 - 62) |
| INTEGRAL | Likelihood = -2.97 | Transmembrane | 156 - 172 (156 - 177) |
| PERIPHERAL | Likelihood = 12.63 | | 284 |

modified ALOM score: 2.92

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 266/382 (69%), Positives = 317/382 (82%)

Query: 10 TIEYIFVLIGAFLLSIILTPPIIRVISLKVGAVDKPNARRINKVPMPSGGLAIFLSFVVT 69
 T++++ VLI L S++LTP++R +L+VGAVD PNARRINKVPMPS+GGLAI +SFV+
 Sbjct: 7 TLKFVLVLIATLLTSLVLTPLVRFALRVGAVDNPNARRINKVPMPSAGGLAIIISFVIA 66

Query: 70 TLFFMPMAASRHFIEVSFYHYILPVIIGGLVVTITGFIIDIFELRPYKMLGIIIAAIII 129
 TL MPM SYF YILPV++G LV+ TGFIDD++EL P+ K LGI++ A+II
 Sbjct: 67 TLALMPMILKTQIGGKSYFEYILPVVLGALVIALTGFIDDVVELSPKIKFLGILLGAVII 126

Query: 130 WKFTFRFDSFKIPIGGPILLEFGPILTFFLTTLWIIISITNAINLIDGLDGLVSGVSIISL 189
 W FT FRFDSFKIP GGP+L F P L+FFLT+LW+++ITNA+NLIDGLDGLVSGVS+ISL
 Sbjct: 127 WIPTDFRFDSEFKIPFGGPMLEHFNPLSFFLTILWVVAITNAVNLIIDGLDGLVSGVSMISL 186

Query: 190 ATMAVVSYYFFLPKIDFFLTITIVILIASIVGFFPYNYHPAIIYLG DAGALFIGFMIGVLS 249
 TM +VSYYFFL D FLTLTI +LI +I GFFPYNYHPAIIYLG DAGALFIGFMISVLS
 Sbjct: 187 TTMGLVSYYFFLYDTIDIFLTITIFVLIFAIAGFFPYNYHPAIIYLGDTGALFIGFMISVLS 246

Query: 250 LQGLKNSTAVAVITPVIIILGVPIIDTAVAIIVRRKLSGKKISEADKMHLHRLLSMGFTHR 309
 LQGLKN+TAVAV+TP+I+LGVPI+DT VAI+RR LSG+K EAD MHLHRL+MGFTHR
 Sbjct: 247 LQGLKNATAVAVVTPPIIVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHRLLAMGFTHR 306

Query: 310 GAVLVVYGGIAIIFSLIALLLNVSSRIGGIFLLALLAMEIFIEGLNIWGENRTPLFNLL 369
 GAVLVVYGGIA+ FSL++LLNVSSR+GGI L++ + A+EIFIEGL IWG RTPLF LL
 Sbjct: 307 GAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAPALEIFIEGLEIWGPKRTPLFRLL 366

Query: 370 KFIGNSDYRQSVIAKYSDKHQK 391
 FIGNSDYRQ V+AKY K +K
 Sbjct: 367 AFIGNSDYRQEVVAKYRRKKKK 388

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5791> which encodes the amino acid sequence <SEQ ID 5792>. Analysis of this protein sequence reveals the following:

-2103-

Possible site: 32

>>> Seems to have an uncleavable N-term signal seq

5 INTEGRAL Likelihood = -8.28 Transmembrane 9 - 25 (1 - 33)
 INTEGRAL Likelihood = -8.17 Transmembrane 201 - 217 (198 - 221)
 INTEGRAL Likelihood = -7.64 Transmembrane 308 - 324 (305 - 329)
 INTEGRAL Likelihood = -7.17 Transmembrane 55 - 71 (51 - 74)
 INTEGRAL Likelihood = -7.06 Transmembrane 145 - 161 (138 - 170)
 INTEGRAL Likelihood = -6.58 Transmembrane 260 - 276 (251 - 278)
 10 INTEGRAL Likelihood = -6.21 Transmembrane 180 - 196 (172 - 198)
 INTEGRAL Likelihood = -5.95 Transmembrane 331 - 347 (330 - 353)
 INTEGRAL Likelihood = -5.68 Transmembrane 87 - 103 (82 - 104)
 INTEGRAL Likelihood = -3.93 Transmembrane 113 - 129 (112 - 133)
 INTEGRAL Likelihood = -2.60 Transmembrane 233 - 249 (232 - 250)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]

Identities = 289/381 (75%), Positives = 334/381 (86%), Gaps = 1/381 (0%)

25 Query: 5 TIDYVLVLIGALLMSLFLTPLVRFLAFRVGAVDNPVNARRVNKVPMPSSGGLAIFMSFLVA 64
 T+ +VLVLI LL SL LTPPLVRF A RVGAVDNPVNARR+NKVPMP++GGLAI +SF++A
 Sbjct: 7 TLKFVVLVIATLLTSLVLTPLVRFALRVGAVDNPVNARRINKVPMPSSAGGLAIIISFVIA 66

30 Query: 65 SLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKLMFGILIGAVI 124
 +L L+P+ K G++YF YILPVV+GA VI LTGF+DD+YELSPK+K GIL+GAVI
 Sbjct: 67 TLALMPMILK-TQIGGKSYFEYILPVVLGALVIALTGFIDDVYELSPKIKFLGILLGAVI 125

35 Query: 125 VWAFTDFKFDSEFKIPFGGPLLVFGPFLTLFLTVLWIVSITNAINLIDGLDGLVSGVSIIS 184
 +W FTDF+FDSEFKIPFGGPL+L F PFL+ FLT+LW+V+ITNA+NLIDGLDGLVSGV+IS
 Sbjct: 126 IWIFTDFRFDSEFKIPFGGPLMLHFNPFLLSFFLTILWVVAITNAVNLIDGLDGLVSGVSMIS 185

40 Query: 185 LVTMAIVSYFFLPQKDFLTLTLVLISAIAGFFPYNYHPAMIYLGDTGALFIGFMIGVL 244
 L TM +VSYFFL D FLTLTI VLI AIAGFFPYNYHPA+IYLGDTGALFIGFMI VL
 Sbjct: 186 LTTMGLVSYFFLYDTDFLTLTLFVLI FAIAGFFPYNYHPAIYLGDTGALFIGFMISVL 245

45 Query: 245 SLQGLKNSTAVAVVTPVILGVPIMDTIVAIIRSLSGQKFYEPDKMHLHRRLLSMGFTH 304
 SLQGLKN+TAVAVVTP+I+LGVPI+DT VAIIR+LSGQKFYE D MHLHRRLL+MGFTH
 Sbjct: 246 SLQGLKNATAVAVVTPIIVLGVPIVDTTVAIIRTLSGQKFYEADNMHLHRRLLAMGFTH 305

50 Query: 305 RGAVLVVYGITMLFSLISLLLVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEKRTPLFNL 364
 RGAVLVVYGI M FSL+SLLLNVSSR+GG+LLM+G+ F LE+FIEGLEIWG KRTPLF L
 Sbjct: 306 RGAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFLEIFIEGLEIWGPKRTPLFRL 365

Query: 365 LKFIGNSDYRQAMLLKWKKEK 385
 L FIGNSDYRQ ++ K++ KK
 50 Sbjct: 366 LAFIGNSDYRQEVVAKYRRKK 386

An alignment of the GAS and GBS proteins is shown below.

Identities = 282/384 (73%), Positives = 334/384 (86%), Gaps = 1/384 (0%)

55 Query: 6 MIPFTIEYIFVLIGAFLLSIILTPPIIRVISLKVGAVDKPNARRINKVPMPSSGGLAIFLS 65
 M FTI+Y+ VLIGA L+S+ LTP++R ++ +VGAVD PNARR+NKVPMP+SGGLAIF+S
 Sbjct: 1 MFSFTIDYVLVLIGALLMSLFLTPLVRFLAFRVGAVDNPVNARRVNKVPMPSSGGLAIFMS 60

60 Query: 66 FVVTTLFFMPMAAS-RHFIEVSYFHYILPVIIGGLVVTTTGFIDDFELRPRYKMLGIII 124
 F+V +L +P+A+ F +YF YILPV+IG V+T TGF+DD++EL P+ KM GI+I
 Sbjct: 61 FLVASLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKLMFGILI 120

65 Query: 125 AAIIWKFTHFRFDSEFKIPGGPLLEFGPIITFFLTVLWIIISITNAINLIDGLDGLVSGV 184
 A+I+W FT F+FDSEFKIP GGPLL FGP LT FLTVLWI+SITNAINLIDGLDGLVSGV
 Sbjct: 121 GAVIVWAFTDFKFDSEFKIPFGGPLLVFGPFLTLFLTVLWIVSITNAINLIDGLDGLVSGV 180

-2104-

Query: 185 SIISLATMAVVSFFLPKIDFFLTITVILIASIVGFFPYNYHPAIIYLG DAGALFIGFM 244
 SIISL TMA+VSFFFLP+ DFFLTITI++LI++I GFFPYNYHPA+IYLG DAGALFIGFM
 Sbjct: 181 SIISLVTMAIVSYFFLPQKDFLTITLVLISAIAGFFPYNYHPAMTYLGD TGALFIGFM 240

5 Query: 245 IGVLSQLGLKNSTAVAVITPVIIILGVPIIDTAVAIIVRRKLSGKKISEADKMHLHRLLSM 304
 IGVLSQLGLKNSTAVAV+TPVIIILGVPI+DT VAI+RR LSG+K E DKMHLHRLLSM
 Sbjct: 241 IGVLSQLGLKNSTAVAVVTPVIIILGVPIIDTVAIIIRSLSGQKFYEPAKMHLHRLLSM 300

10 Query: 305 GFTHRGAVLVVYGIATIFSLIALLLNVSSRIGGIFLLALLLAMEIFIEGLNIWGENRTP 364
 GFTHRGAVLVVYGI ++FSLI+LLLNVSRRIGG+ L+L LL +E+FIEGL IWGE RTP
 Sbjct: 301 GFTHRGAVLVVYGITMLFSLISLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEK RTP 360

Query: 365 LFNLLKFIGNSDYRQSVIAKYSK 388
 LFNLLKFIGNSDYRQ+++ K+ +K
 15 Sbjct: 361 LFNLLKFIGNSDYRQAMLLKWK 384

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1863

20 A DNA sequence (GBSx1970) was identified in *S. agalactiae* <SEQ ID 5793> which encodes the amino acid sequence <SEQ ID 5794>. This protein is predicted to be negative regulator of genetic competence. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3460(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9483> which encodes amino acid sequence <SEQ ID 9484> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA82113 GB:AB022909 negative regulator of genetic competence
 [Streptococcus mutans]
 35 Identities = 168/248 (67%), Positives = 205/248 (81%), Gaps = 9/248 (3%)

Query: 1 MEMKQISETTLKITISMEDLEDGMELKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
 MEMKQISETTLKITISMEDLE+RGMELKDFLIPQEKTEEFFY+VMDELDLPENFK SGML
 40 Sbjct: 1 MEMKQISETTLKITISMEDLEERGMEKDFLIPQEKTEEFFYTVMDELDLPENFKSGML 60

Query: 61 SFRVTPKKDRIDVFTKSELSKDLNLEELADLGDISKMSPEDEFFKTLEQSMLEKGD TDAH 120
 SFRVTP+ DRIDVFTKSE++K+LNLE+L+D DISKMSPEDEFF TLE++M EKGD A
 45 Sbjct: 61 SFRVTPRNDRIDVFTKSEINKNLNLEDLSDFDDISKMSPEDEFFNTLEETMREKGDAAAL 120

Query: 121 AKLAEIENMMDKATQEVVEENVSEEQPEKEVETIGYVHYVDFDNIEAVVRFSQTIDFPI 180
 KLAEIE ++ TQ+ E+ ++E+ + YVH+V DF NI+ V+ F++T+D+ +
 50 Sbjct: 121 DKLAEIEKREEEKTQQ--EKGETKEKRD-----YVHFVLD FNIQQVVISFAKTVDYDV 171

Query: 181 EASELYKNGKGYHMTILLDLNQPSYFANLMYARMLEHANVGT KTRAYLKEHSIQLIHDD 240
 EASEL+K YHMT+LL+LE++P Y+A+LM+ARMLHA GTKTRAYL EH +QLI D
 55 Sbjct: 172 EASELFKESDAYHMTVLLNLEDKPDYYADLMFARMLEHAGRGTKTRAYLLEHGVQLIKAD 231

Query: 241 AISKLQMI 248
 A+ +LQMI
 60 Sbjct: 232 ALQELQMI 239

-2105-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5795> which encodes the amino acid sequence <SEQ ID 5796>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/253 (67%), Positives = 209/253 (82%), Gaps = 2/253 (0%)

```

Query: 1  MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
          MEMKQISETTLKITISM+DLE+RCMELKDFLIPQEKTEEFFYSVMDELDLP+NFK+SGML
Sbjct: 3  MEMKQISETTLKITISMDDLEERGMEKDFLIPQEKTEEFFYSVMDELDLPDNFKDSGML 62

Query: 61  SFRVTPKKDRIDVFVTKSELSKDNLNLEELADLGDISKMSPEDFFKTLEQSMLEKGDTHAH 120
          SFRVTP+KDR+DVFTKSE++KD+NLE+LA+ GD+S+M+PEDFFK+LEQSM EKGD AH
Sbjct: 63  SFRVTPRKDRLDVFVTKSEINKDINLEDLAIEFGDMSQMTPEDFFKSLEQSMREKGDVKAH 122

Query: 121 AKLAEIENMMDKATQEVV--EENVSEEQPEKEVETIGYVHYVDFDNIIEAVVRFSTIDF 178
          KL +IE +M+ + + + ++ E E + YVHYV DF I V F++TIDF
Sbjct: 123 EKLEKIEEIMEDVVEATLANQSEAADPSTNHESEPLDVHYVLDFTITEAVAFKTIIDF 182

Query: 179 PIEASELYKNGKGYHMTIILLDLENQPSYFANLMYARMLEHANVGTCTRAYLKEHSIQLIH 238
          IEASELYK YHMTIILLD++ QPSYFAN+MYAR++EHAN G+KTRAYL+EH +QL+
Sbjct: 183 SIEASELYKGSNCYHMTIILLDVQQPSYFANVMYARLIEHANPGSKTRAYLQEHGLQLML 242

Query: 239 DDAISKLQMIEMG 251
          D A+ +LQ IE+G
Sbjct: 243 DGAVEQLQKIELG 255

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1864

A DNA sequence (GBSx1971) was identified in *S.galactiae* <SEQ ID 5797> which encodes the amino acid sequence <SEQ ID 5798>. This protein is predicted to be BacA (bacA). Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.02 Transmembrane 115 - 131 (111 - 135)

INTEGRAL Likelihood = -8.97 Transmembrane 227 - 243 (219 - 247)

INTEGRAL Likelihood = -7.86 Transmembrane 48 - 64 (44 - 69)

INTEGRAL Likelihood = -7.27 Transmembrane 263 - 279 (260 - 279)

INTEGRAL Likelihood = -7.22 Transmembrane 87 - 103 (85 - 107)

INTEGRAL Likelihood = -3.50 Transmembrane 2 - 18 (1 - 19)

----- Final Results -----

bacterial membrane --- Certainty=0.4609(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]

Identities = 101/275 (36%), Positives = 165/275 (59%), Gaps = 22/275 (8%)

Query: 7 LKALFLGVVEGVTEWLPVSSSTGHLLLVQEFMKLNQSKSFVEMFNIVIQLGAIMAVIVIYF 66

-2106-

L+A+ L V+EG+TE+LPVSSTGH+I+ F + + F ++F IVIQLGAI++V+V+YF
 Sbjct: 4 LQAIVLAVIEGITEFLPVSSSTGHMIIASSFFGIAH-EDFTKLFTIVIQLGAILSVVVLYF 62
 Query: 67 KRLNPFQPGKSAREIRLTWQLWLKVVIACIPISILIALPFDNWFEAHFNFMIPIAIALIFY 126
 KR FQ T + K+++A IP++++ L ++ + + +A++L+
 Sbjct: 63 KRF--FQ-----TLDYFYLKLLVAFIPAVVLGLLLSDFIDGLLENPVTVAVSLIG 110
 Query: 127 GFVFI---WVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGAIL 182
 G + + W NA Q ++Y A IG FQ +++PG SRSGA+I+G +
 Sbjct: 111 GLILLKVDWFWNNPNAETSQ-----KITYLQALKIGLFQCIAMIPGVSRSGASIVGMS 165
 Query: 183 IGTSRSVAADFTFFLAIPTMFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSILYV 242
 SR+ AA+F+FFLA+PTM G + K Y+ G LS DQ IL++ ++ AF+V+L
 Sbjct: 166 QKLSRTTAAEFSPFLAVPTMLGATVKKCYDYKAGFELSHDQVNILIIIGNVAFIVALLA 225
 Query: 243 IRFLTIDYVKRHDFITFGKYRIVLGSLLILYWLTVH 277
 I+ ++ ++ F +FG YRI+ G +L+L +H
 Sbjct: 226 IKTFISFLTNGFKVFGYRIIAGIILLIHHFFIH 260

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5799> which encodes the amino acid sequence <SEQ ID 5800>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.30 Transmembrane 225 - 241 (219 - 247)
 INTEGRAL Likelihood = -9.24 Transmembrane 115 - 131 (109 - 135)
 INTEGRAL Likelihood = -7.64 Transmembrane 48 - 64 (44 - 69)
 INTEGRAL Likelihood = -7.43 Transmembrane 87 - 103 (85 - 108)
 INTEGRAL Likelihood = -5.15 Transmembrane 263 - 279 (262 - 279)
 INTEGRAL Likelihood = -3.82 Transmembrane 2 - 18 (1 - 19)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 102/269 (37%), Positives = 169/269 (61%), Gaps = 14/269 (5%)
 Query: 7 LKAIFFGIIEGITEWLPVSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIVMLIYF 66
 L+AI +IEGITE+LPVSSTGH+I+ F + + F ++F IVIQLGAI++V++YF
 Sbjct: 4 LQAIVLAVIEGITEFLPVSSSTGHMIIASSFFGIAHED-FTKLFTIVIQLGAILSVVVLYF 62
 Query: 67 ERLNPFQPGKTAREVQLTWQLWLKVVIACIPISILIAVPLDNWFEAHFYFMVPIAIALIVY 126
 +R FQ T + K+++A IP++++ + L ++ + V +A++L++
 Sbjct: 63 KRF--FQ-----TLDYFYLKLLVAFIPAVVLGLLLSDFIDGLLENPVTVAVSLIG 110
 Query: 127 GIAFIWIEKRNAQQEPAVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGAILGTS 186
 G+ + +++ A T +++Y A IG FQ +++PG SRSGA+I+G + S
 Sbjct: 111 GLILLKVDWFWNNPNAETS-QKITYLQALKIGLFQCIAMIPGVSRSGASIVGMSQKLS 169
 Query: 187 RTVAADFTFFLAIPTMFGYSGLKAVKFFLDGHHLDFQAQVILLVASLTAFVVSLLAIRFL 246
 RT AA+F+FFLA+PTM G + K ++ G L QV IL++ ++ AF+V+LLAI+
 Sbjct: 170 RTTAAEFSPFLAVPTMLGATVKKCYDYKAGFELSHDQVNILIIIGNVAFIVALLAIKTF 229
 Query: 247 TDYVKKHDFITFGKYRIVLGSLLLIYSFF 275
 ++ K+ F +FG YRI+ G +LL+ FF
 Sbjct: 230 ISFLTNGFKVFGYRIIAGIILLIHHFF 258

60 An alignment of the GAS and GBS proteins is shown below.

Identities = 227/272 (83%), Positives = 253/272 (92%)

Query: 1 MLIIELLKALFLGVVEGVTEWLPVSSTGHLILVQEFMKNQSKSFVEMFNIVIQLGAIMA 60
 MLIIELLKA+F G++EG+TEWLPVSSTGHLILVQEF++LNQ K+F+EMFNIVIQLGAI+A

-2107-

Sbjct: 1 MLIIELLKAIFFGIIEGITEWLPVSSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIA 60

Query: 61 VIVIFYKRLNPFQPGKSAREIRLTWQLWLKVVIACIPISILIALPFDNWFEAHFNFMPIA 120
V++IYF+RLNPFQPGK+ARE++LTWQLWLKVVIACIPISILIA+P DNWFEAHF FM+PIA

5 Sbjct: 61 VMLIYFERLNPFQPGKTAREVQLTWQLWLKVVIACIPISILIAVPLDNWFEAHFYFMVPIA 120

Query: 121 IALIFYGFFVIWVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGA 180
IALI YG FIW+EKRNA +P VTELA MSYKTAFFIGCFQVLSIVPGTSRSGATILGA

10 Sbjct: 121 IALIVYGIAFIWIEKRNAQQEPAVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGA 180

Query: 181 IIIGTSRSVAADFTFFFLAIPMTFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSL 240
II+GTSR+VAADFTFFFLAIPMTFGYSGLKAVK+FLDG+ L Q LILLVASLTAFVVSL

Sbjct: 181 IILGTSRTVAADFTFFFLAIPMTFGYSGLKAVKFFLDGHHLDFQVLILLVASLTAFVVSL 240

15 Query: 241 YVIRFLTDYVKRHDFITFGKYRIVLGSLLILY 272
IRFLTDYVK+HDFITFGKYRIVLGSLL++Y

Sbjct: 241 LAIRFLTDYVKKHDFITFGKYRIVLGSLLLIY 272

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 1865

A DNA sequence (GBSx1972) was identified in *S.galactiae* <SEQ ID 5801> which encodes the amino acid sequence <SEQ ID 5802>. Analysis of this protein sequence reveals the following:

Possible site: 42

25 >>> Seems to have no N-terminal signal sequence

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.65 | Transmembrane | 494 - 510 (488 - 519) |
| INTEGRAL | Likelihood = -8.01 | Transmembrane | 263 - 279 (256 - 288) |
| INTEGRAL | Likelihood = -5.95 | Transmembrane | 25 - 41 (20 - 43) |
| INTEGRAL | Likelihood = -4.94 | Transmembrane | 475 - 491 (473 - 493) |

30 ----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.4461(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

35

A related GBS nucleic acid sequence <SEQ ID 9481> which encodes amino acid sequence <SEQ ID 9482> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
[Methanococcus jannaschii]
Identities = 41/172 (23%), Positives = 78/172 (44%), Gaps = 19/172 (11%)

Query: 479 LISFVVIITLFLNYFTYFCIYLLLFVILLLNKIIIFMMTRKISNGYIVTEDGASRVYQW 538
+IS ++ ++ F+ ++ + ++ ++ II +T G ++ +W

45 Sbjct: 442 VISILLAVFLYFIPKYSQTFNEVFYLSIVFVQNIIILALTPTSIFGRWKANYEKEKL-EW 500

Query: 539 TSFRNMLRDIKSFDRSELESIVLWNRILVYATLFGYADRVEKALR-VNQIDIPERFANID 597
+F+N L ++ + E I +W L+Y T G D+V +A++ +N ++ + I

50 Sbjct: 501 DAFKNFLSNLAMIKKYSPEDISIWKDWLTYGTALGVGDKVVEAMKSLNLSLVADYVIIH 560

Query: 598 SHQFAISVNQSSNHFSTITEDVSHASNFSVNSGGSSGGFSGGGG--GGGGGA 647
S+ ++ + S+ ST GS GGF GGG GGGGGA

Sbjct: 561 SNYDSMKTSDSVYSSTT-----GSGGGFGAGGGFGGGGGGA 597

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5803> which encodes the amino acid sequence <SEQ ID 5804>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have a cleavable N-term signal seq.

-2108-

INTEGRAL Likelihood = -7.91 Transmembrane 486 - 502 (483 - 508)
 INTEGRAL Likelihood = -5.89 Transmembrane 465 - 481 (460 - 483)
 INTEGRAL Likelihood = -2.18 Transmembrane 244 - 260 (241 - 260)

5

----- Final Results -----

bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10

The protein has homology with the following sequences in the databases:

>GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577

[Methanococcus jannaschii]

Identities = 59/263 (22%), Positives = 106/263 (39%), Gaps = 14/263 (5%)

15

Query: 369 FLDMAFGNKVTLFVDQLFSQYHYDADTIKQLKKTYKGKKLEQEVRSSEQVIKAMKKASA 428

++ + G K + + L + Y++D +K L K K + E +S Q K+ K

Sbjct: 346 YIKIMNGGKIEILKTDLENLDVYESDVMKFLMKYSKNNVDFPEYIKSLAQKYKSSKDKLK 405

20

Query: 429 AITNNVLETIKKLNLPDTPYQMTA--EKRKSNSVQGLGCLLLILNSGLLIYLAIKESGL 486

+ + E K + P ++ A E R + L + ++L L ++

Sbjct: 406 KDKD---ELDKIMEYPRYSSKVVNAFLETRGKKIIIALLVISILLAVFLYFIPKYSQTFN 462

Query: 487 ALIYLALMVLTMCLGFYISLKLQYKKGIIETPEGGVRLHQWQSFKNMIRDIDKFEDVAI 546

+ YL+++ + I L L G +W +FKN + ++ + +

25

Sbjct: 463 EVFYLIVFVQ---NIILALTPTSLFGRWKANYKEKLEWDAFKNFLSNLAMIKKYSP 518

Query: 547 EGLVVMNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYATPTFVSSL 606

E + +W L+Y T G KV +K ++ + V + Y + T V S+

Sbjct: 519 EDISIWKDLIYGTALGVGDVVEAMKSLNLS-----ELVADYVVIHSNYSMTSVDSV 573

30

Query: 607 SSATTSSNFSVSSGGGSGGGGG 629

S+TT S +GGG GGGGG

Sbjct: 574 YSSTTSGGGGFGAGGGFGGGGG 596

35

An alignment of the GAS and GBS proteins is shown below.

Identities = 241/635 (37%), Positives = 372/635 (57%), Gaps = 18/635 (2%)

Query: 22 MKKCFLAICLALSFFMVSVQADEVDYNIPIHYEGNLTIHNDNSADFTKEKVTYQFDSSYNGQ 81

MKK + + L S + ++A +VDY+I +YEG L + +N+A F +KVTYQFD+SYNGQ

40

Sbjct: 1 MKKILMTLVLCFSLGIRIKAADVDYSITNYEQQLLSKENTARFEQKVTYQFDTSYNGQ 60

Query: 82 YVTLGTAGKLDPDNFDINNKQVEVSINGKVRKVSQIEDLEDGYRLKVFNGGEAGDTVKV 141

Y++LG G LP F I+ KP+VEV NG+ VS + DL DGYRLK++N G+AGD V V

Sbjct: 61 YISLGRTHLPAGFAIDQKPKVEVYQNGQVPVSQEFSDLDGYRLKLYNAGQAGDKVDV 120

45

Query: 142 NVQWKLKNVLFMHKDVGELNWIPIISDWDKTLEKVDWFISTDKKVALSRLWGHLYL-KTP 200

V W+L ++L ++DV ELNW PISDWDKTLEKV ++T + S LW H GY K P

Sbjct: 121 KVIWQLHLLTAYQDVAELNWTPIISDWDKTLEKVSILTPTPTDIQDSNLWAHGGYQKKP 180

50

Query: 201 PKIRQNNRYHLTAFAFNVRKLEFHYGWDYSYF--NLPTNSKNYKKIEYQEKMIERHGF 258

+++ N+RY + A NV+ +LE H YWD+ P + + K KI E I R

Sbjct: 181 QVLKEGNSRYQINAKNVSCQLELHAYWDKKALLGKEPVDVSTSKKNKIVALETKISRRRT 240

Query: 259 ILSFLLRILLPSFFIIVTLFISIRVFLFRKKVKNYGFQFPKEHHLYEAPEDLSPLELTQSI 318

+L L ++P + L+ I+ +K+ N+Y H YE PEDLSPL LTQ+I

55

Sbjct: 241 LLQLLFGKVIPLVEVGFLWLQLIQFTRLKKQFNRYHLNHTDHSYEVPEPDLSPVLVTQAI 300

Query: 319 YSMSFKNFQ---DEEKKTHL---ISQEQLIQSILLDLIDRKVL---NYDDNLLSLANLD 368

Y SF E +K + ++ E L+Q+ LLDLID+KVL L ++ LD

60

Sbjct: 301 YGQSFAYLSPTASESQKLLIPKGVTFEALVQATLLDLIDQKVLTLTKEEGKAYLEISQLD 360

Query: 369 RASDAEIDFIEFAFADSTSLKPDQLFSNYQFSYKETLRELKKQHKASDLQTQMRRRGSNA 428

R +D E F++ AF + +L DQLFS Y + +T+++LKK +K L+ ++R+

Sbjct: 361 RVTDEEAFLDMAFGNKVTLFVDQLFSQYHYD-ADTIKQLKKTYKGKKLEQEVRSSEQV 419

65

Query: 429 LSRITRLTRLISKDNINSLRRKGISSPYRKMSSEESKELSRKRFSYLSPLISFVVIYT 488

-2109-

```

      + + + + I+ + + + + + + YR+M+ E ++ + ++ L +++ ++IY
Sbjct: 420 IKAMKKASAAITNNVLETIKLNLPTYRQMTFAEKRSNSVQGLGCLLLILNSGLLIY- 478

Query: 489 LFLNYFTYFCIYLLLFVILLNKKIIFMMTRKISNGYIVTEDGASRVYQWTSFRNMLRDI 548
      L +          IYL L + + L I + + I T +G R++QW SF+NM+RDI
Sbjct: 479 LAIKESGLALIYALMVLTMCLGFYISLKLQYKKLGIETPEGGVRLHQWQSFKNMIRDI 538

Query: 549 KSFDRSELESIVLWNRILVYATLFGYADRVEKALRVNQIDIPERFANIDSHQFAISVNQS 608
      F+ +E +V+WNR+LVYATLFGYA +VE+ L+V++I +PE + + + ++ + +
Sbjct: 539 DKFEDVAIEGLVWNRVLRVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYAT 598

Query: 609 SNHFSTITEDVSHASNFSVNSGGSSGGFSGGGGGG 643
      + F + + +SNFSV+SG GG SGGGGGG
Sbjct: 599 TPTFVSSLSSATTSSNFSVSSG--GGISGGGGGG 630

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8921> and protein <SEQ ID 8922> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 4
McG: Discrim Score:      10.29
GvH: Signal Score (-7.5): 3.11
      Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 3 value: -8.65 threshold: 0.0
      INTEGRAL    Likelihood = -8.65  Transmembrane  475 - 491 ( 469 - 500)
      INTEGRAL    Likelihood = -8.01  Transmembrane  244 - 260 ( 237 - 269)
      INTEGRAL    Likelihood = -4.94  Transmembrane  456 - 472 ( 454 - 474)
      PERIPHERAL  Likelihood = 2.28    540
modified ALOM score: 2.23

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no homology with any sequences in the databases.

Example 1866

A DNA sequence (GBSx1973) was identified in *S.agalactiae* <SEQ ID 5805> which encodes the amino acid sequence <SEQ ID 5806>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -8.86  Transmembrane  301 - 317 ( 295 - 324)
      INTEGRAL    Likelihood = -6.05  Transmembrane  479 - 495 ( 473 - 496)
      INTEGRAL    Likelihood = -0.59  Transmembrane  369 - 385 ( 369 - 385)

----- Final Results -----
      bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
      [Synechocystis sp.]
Identities = 147/534 (27%), Positives = 256/534 (47%), Gaps = 75/534 (14%)

```

-2110-

Query: 4 ILLSLFTALLITFGGMTSIQADEYLRVGMEEAAYAPFNWNTQNDNTNGAVPIEGTDQYANGY 63
 +LL++ LL F ++ + + V E + PF T E T Q G+
 Sbjct: 24 VLLAIAIPLLPAFSQVSR---QTIIVATEPTFPFFEMTD-----EATGQLT-GF 68

Query: 64 DVQVAKKLAKLKNKKVVVVKTKWEGLVPALTSGKLDMIAGMSPTTEERKKEINFSPYYI 123
 DV + + + + V + ++G++PAL S + I+ ++ T ER + ++FS PY+
 Sbjct: 69 DVDLIQAIGEAAQVTVDIQGYPFDDGIIPALQSNVTGAAISAITITPERAQSVSFSSPYFK 128

Query: 124 SEPTLVVNAEGKYTNAKNISDFKNKVTAQQGVVLYNLIDQINGVKKEVAMGDFNQLRQA 183
 S L + + KN+ D + ++ G + + G K + +F+ + A
 Sbjct: 129 S--VLAIAVQDGNDTIKNLKDLEGKRLAVAIGTTGAMVATNVPGAK---VINFDSTSITA 182

Query: 184 VE---SGVVDAYVSRPDATSAQTANPKLKMIELHQFCKTSDADTNISVGMKRGDNRIHQ 240
 ++ +G DA +++RP A + L+ +++ + D I++ + INQ
 Sbjct: 183 LQELVNGNADAVINDRPVLLYA- IKDAGLRNVKISADVGSSEY-YGIAMPLAP-PGEINQ 239

Query: 241 VNQVL-----ESISRDQIALMDKMIKEQ-----PSV-----KKEKNGK 274
 +VL + I A+ +K E+ PS+ + + N
 Sbjct: 240 TREVLNQGLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVGKVGTAQSLTERSQANPN 299

Query: 275 PNFFEQMATILKNNGSQFLRGTTATLLISMVGTIVGLFIGLLIGVFTAPKSDNKLKAAL 334
 NF + T+ +N +G+ T+L++ GL G + + A SD
 Sbjct: 300 DNF---LITLFRN---LFKGSILTVLLTAFSVFFGLIGGTGVAI---ALISD----- 342

Query: 335 QKLLGWLLNIYIEVFRGTPMIVQSMVIYGTAAQF-----GVSLDRTLAAIFIVSINTGA 389
 K L + IY+E FRGTPM+VQ +IY+G F G+++DR AAI +S+N A
 Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLITIDRFPAIIALSLNVAA 402

Query: 390 YMSEIVRGGIFSVDKGQFEAATALGFTHGQTMKIVLPQVVRNIPATGNEFVINIKDTS 449
 Y++EI+RGGI S+D+GQ+EA +LG + QTM++++ PQ R ILP GNEF+ IKDTS
 Sbjct: 403 YLAEIIRGGIQSIDQGWEEACSLGMSFWQTMKEVIFPQAFRRILPPLGNEFITLIKDTS 462

Query: 450 VLNVISVVELYFSGNTVATQTYQYFQFTIIAIYFILFTVTRILRYIEKRFD 503
 + VI EL+ G + TY+ F+ + +A++Y +LT + + +++E D
 Sbjct: 463 LTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLTTTSSFFVKWLENYMD 516

There is also homology to SEQ ID 1194.

A related GBS gene <SEQ ID 8923> and protein <SEQ ID 8924> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 McG: Discrim Score: 6.23
 GvH: Signal Score (-7.5): 0.11
 Possible site: 24

>>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 3 value: -8.86 threshold: 0.0

| | | | |
|------------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.86 | Transmembrane | 301 - 317 (295 - 324) |
| INTEGRAL | Likelihood = -6.05 | Transmembrane | 479 - 495 (473 - 496) |
| PERIPHERAL | Likelihood = 1.32 | | 441 |

modified ALOM score: 2.27

*** Reasoning Step: 3

----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.4545(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

The protein has homology with the following sequences in the databases:

34.3/57.3% over 462aa

Synechocystis PCC6803

EGAD|48193| glutamine-binding periplasmic protein/glutamine transport system permease protein Insert characterized

GP|1652664|dbj|BAA17584.1||D90907 glutamine-binding periplasmic protein {Synechocystis sp.} Insert characterized

PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803) Insert characterized

ORF01242(454 - 1809 of 2148)
EGAD|48193|sl11270(54 - 516 of 530) glutamine-binding periplasmic protein/glutamine
transport system permease protein {Synechocystis PCC6803}GP|1652664|dbj|BAA17584.1||D90907
glutamine-binding periplasmic protein {Synechocystis sp.}PIR|S77250|S77250 hypothetical
protein - Synechocystis sp. (strain PCC 6803)
%Match = 12.3
%Identity = 34.2 %Similarity = 57.2
Matches = 128 Mismatches = 149 Conservative Sub.s = 86

10 20 30 40

444 474 495 525 555 585 615 645
FNWTTQNDNTNGAVPIEGTDQ--YANGYDQVAKKLAKLKNKKVVVVVKTKEGLVPALTSGKLDMIAGMSPTEERKKEI
 | | ||: | :| : : : : | : : :| :|| | : | : : | | : :
RQTIIVATEPTFPFPEMTDEATGQLTGFDVLDLIQAIGEAQVTVDIQGYFPDGIIPALQSNVTGA AISAITITPERAQS
 50 60 70 80 90 100 110 120

675 705 735 765 795 825 855 885
NFSKPYIISEPTLVVNAEGKYTNAKNISDFKNQKVTAQQGVLYNLIDQINGVKKEVAMGDFNQLRQAVESGVVDAYVSE
: || | |: : | : | | : | ::
SFSSPFYFKSVLAIQV-DGNDT-IKNLKDLLEGKRLLAVAIGTTGAMVATNVPGAKVTNFDSITSALQELV-NGNDAVIN

130 140 150 160 170 180 190

903 957 987
 RP-----DATSAQTANPKLK-MIELHQG-FKTSADTNISV
 || | || | : | : || |
 RPVLLYAIKDGALRNVKISADV~NPPFLPLVAPSLVGKVGTAQSLTERSQANPNDFLITLFRNLFKGS-----
 210 270 280 290 300 310

```

1017      1047      1077      1107      1137      1167      1197      1227
GMRKGDNRINQVNVQVLESISRDKQIALMDKMIKEQPSVKKEKNGKPNFFEQMATILKNNGSQFLRGTAITLLISMVGTV
                                     :::: |
-----|ITVLLTAF
                                     320

```

1257 1284 1314 1344 1374 1404 1419 1449
GLFIGLLIGV-FRTAPKSDNKLKAALQKLLGWLLNIYIEVFRGTPMIVQSMVIYGTAAAF----GVSLDRTLAAIFIV
:| ||: | || | | : | : | |||||:| | :||:| | | :|| | || | :
SVFFGLIGGTGVAIALISD-----IKPLQLIRIYVEFFRGTPMLVQLFIYFGLPALFKEIGLGITIDRFPAATIAL
 340 350 360 370 380 390

1479 1509 1539 1569 1599 1629 1659 1689
SINTGAYMSEIVRGGISVDKGQFEAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVLNVISVVELYFSG
|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
SLNVAAYLAEIIRGGIQSIDQGWEEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSLTAVIGFQELFREG
410 420 430 440 450 460 470

1719 1749 1779 1809 1839 1869 1899 1929
 NTVATQTYYQYFQTITIIAIYFILFTVTRILRYIEKRFSDSNYTTGANQLQV*EVGMTQAILEIKHLKKSYSGSNEVLKD
 : ||: | : : | : : || : : : : | | :
 QLIVATTYRAFEVYIAVALVYLLLTTSISFVFKWLENYMDPIGRAKKKAKAATA
 490 500 510 520 530

There is also homology to SEQ ID 5804.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2112-

Example 1867

A DNA sequence (GBSx1974) was identified in *S.agalactiae* <SEQ ID 5807> which encodes the amino acid sequence <SEQ ID 5808>. This protein is predicted to be ATP-binding. Analysis of this protein sequence reveals the following:

```

5      Possible site: 44
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3208(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAB73160 GB:AL139076 putative glutamine transport ATP-binding
      protein [Campylobacter jejuni]
      Identities = 132/241 (54%), Positives = 178/241 (73%), Gaps = 1/241 (0%)

      Query: 5   ILEIKHLKKSYSNEVLKDISLSVNKGEVISIIGSSSGSKSTFLRSINLLEEPSGGEILY 64
      ++E+K+L+K YG EVLK+I+ +++KG+VI+IIG SG GKSTFLR IN LE GEIL
20      Sbjct: 1   MIEVKNLQKKYGELEVLKNINTISKGDVIAIIGPSGGGKSTFLRCINRLELADSGEILI 60

      Query: 65   HGHNVLEKGYDLNNYREKLGVMVFQSFNLFENLNILENAIVAQTTLKRRERQEAKEIAKEN 124
      + N+L+K D+N R+K+ MVPQ FNL F N N++EN + ++EA K AK
25      Sbjct: 61   NKQNILDKEIDINKIRQKYSVMVFQHFNL FANKNV MENLCLTPIKTGILSQEEAIKKAKLL 120

      Query: 125  LNAVGMTEQYWKAKPKQLSGGQKQRVAIARALSVNPEAILFDEPTSALDPEMVGEVLKTM 184
      L VG+ ++ P +LSGGQKQ+AIAR+L +NP+ ILFDEPTSALDPEM+GEVL M
30      Sbjct: 121 LAKVGLADKE-NIMPHKLSGGQKQRIARSLMNPDVILFDEPTSALDPEMIGEVLSIM 179

      Query: 185  QDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIAEQGTKQLFENPTQERTKEFLQRF 245
      +D+AK GLTM++VTHEM FA+ V++R+ FMDKG IA +PK++FENP+ ER +EFL + L
35      Sbjct: 180 KDVAKGLTMLVVTHEMGFARNVANRIFMDKGKIAVDASPKVFENPSNERLREFLNKVL 240

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2157> which encodes the amino acid sequence <SEQ ID 2158>. Analysis of this protein sequence reveals the following:

```

35      Possible site: 27
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.1170(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

45      Identities = 212/246 (86%), Positives = 237/246 (96%)

      Query: 1   MTQAILEIKHLKKSYSNEVLKDISLSVNKGEVISIIGSSSGSKSTFLRSINLLEEPSGG 60
      M+ +I+EIK+LKKSYSNEVLKDISLSVNKGEVISIIGSSSGSKST LRSINLLEEPS G
50      Sbjct: 24   MSNSIIEIKNLKKSYSNEVLKDISLSVNKGEVISIIGSSSGSKSTLLRSINLLEEPSAG 83

      Query: 61   EILYHGHNVLEKGYDLNNYREKLGVMVFQSFNLFENLNILENAIVAQTTLKRRERQEAKEI 120
      +IL+HG +VL + Y+L +YREKLGVMVFQSFNLFENLN+LENAIVAQTTLKRR+R +AE+I
55      Sbjct: 84   QILFHGEDVLAEHYNLTHYREKLGVMVFQSFNLFENLVLENAIVAQTTLKRDRAQAEQI 143

      Query: 121  AKENLNAVGMTEQYWKAKPKQLSGGQKQRVAIARALSVNPEAILFDEPTSALDPEMVGEV 180
      AKENLNAVGMTEQYW+AKPKQLSGGQKQRVAIARALSVNPEA+LFDEPTSALDPEMVGEV
60      Sbjct: 144  AKENLNAVGMTEQYWQAKPKQLSGGQKQRVAIARALSVNPEAMLFDEPTSALDPEMVGEV 203

      Query: 181  LKTMQDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIAEQGTKQLFENPTQERTKEF 240
      LKTMQDLAKSGLTMIIVTHEMEFA++VSDR+IFMDKG+I E+G+P+Q+FENPTQ+RTKEF
      Sbjct: 204  LKTMQDLAKSGLTMIIVTHEMEFARDVSDRIIFMDKGLITEEGSPQIIFENPTQDRTEF 263

```


-2113-

Query: 241 LQRFLK 246
 LQRFLK
 Sbjct: 264 LQRFLK 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1868

A DNA sequence (GBSx1976) was identified in *S. agalactiae* <SEQ ID 5809> which encodes the amino acid sequence <SEQ ID 5810>. This protein is predicted to be hypersensitive-induced response protein. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -17.94 Transmembrane 4 - 20 (1 - 28)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.8175(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9479> which encodes amino acid sequence <SEQ ID 9480> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
 [Zea mays]
 Identities = 127/275 (46%), Positives = 174/275 (63%), Gaps = 1/275 (0%)
 Query: 19 ITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQLRLQSEIIVETKTK 78
 I L V Q T V A I E F G K + + G H + I A + L R + Q + + E T K T K
 Sbjct: 4 ILGLVQVDQSTVAIKENFGKFSEVLEPGCHFLPWCIGQQIAGYLSLRVRLDVRCECTKTK 63
 Query: 79 DNVFVTLNIATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSVPKLTLDLFEKKDE 138
 DNVFVT + QYR + +DA+YKL QI+SY+ D +R++VPKL LD+ FE+K+E
 Sbjct: 64 DNVFVTVVASVQYRALADKASDAFYKLSNTREIQSYVFDVIRATVPKLGLDDAFEQKNE 123
 Query: 139 IALEVQHQAEEEMSTYGYIIVKTLITKVEPD AEVKQSMNEINAAQRKRVAQAELANADKI 198
 I A V + + + M S T Y G Y I V + T L I + E P D V K + + M N E I N A A R V A A E A A + K I
 Sbjct: 124 IAKAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKI 183
 Query: 199 KIVTAAEAEAEKDRHLHGVGIAQQRKAIVDGLADSIQELKDANVTLTTEEQIMSILLTNQYL 258
 + A E E A E L G V G I A + Q R + A I V D G L D S + + T + I M + + L Q Y
 Sbjct: 184 LQIKKAEGEAESKYLAVGVIARQQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYF 243
 Query: 259 DTLNTF-AINGNQTIFLPNNPEGVEDIRTOVLSAL 292
 D T + A + + + F + P + P V + D + Q + L
 Sbjct: 244 DTMREIGASSKSSSVFIPHGPGAVKDVSAQIRDGL 278

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5811> which encodes the amino acid sequence <SEQ ID 5812>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.06 Transmembrane 5 - 21 (1 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2114-

The protein has homology with the following sequences in the databases:

>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
[Zea mays]

Identities = 126/273 (46%), Positives = 174/273 (63%), Gaps = 3/273 (1%)

Query: 23 LYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGI-DKIAARVQLRLQSEIIVETKTKDN 81
L V Q +VAI E FG++ + G H LP+ I +IA + LR+ Q ++ ETKTKDN
Sbjct: 7 LVQVDQSTVAIKENFGKFSEVLEPGCHF-LPWCIGQQIAGYLSLRVQLDVRCEETKTKDN 65

Query: 82 VFVTLNVATQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSVPKLTLDLFEKKDEIA 141
VFVT+ + QYR +DA+YKL QI+SY+ D +R++VPKL LD+ FE+K+EIA
Sbjct: 66 VFVTVVASVQYRALADKASDAFYKLSNTRREQIQSYVFDVIRATVPKLGDDAFEQKNEIA 125

Query: 142 LEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAQELANADKIKI 201
V+ ++ + MSTYGY IV+TLI +EPD VK++MNEINAA R RVAA E A A+KI
Sbjct: 126 KAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKILQ 185

Query: 202 VTAAEAEAEKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLNNEEQIMSILLTNQYLDT 261
+ AE EAE L GVGIA+QR+AIVDGL +S+ E + IM ++L QY DT
Sbjct: 186 IKKAEGEAESKYLAVGIARQQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYFDT 245

Query: 262 LNTFAAKG-NQTLFLPNTPSGVEDIRTQVLSAL 293
+ A + ++F+P+ P V+D+ Q+ L
Sbjct: 246 MREIGASSKSSSVFIPHGPGAVKDVSAQIRDGL 278

An alignment of the GAS and GBS proteins is shown below.

Identities = 254/291 (87%), Positives = 278/291 (95%)

Query: 5 IILTIVLVIVLITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQL 64
I + +++++ ++ ++LYVV+QQ+VAI+ERFG+YQKTATSGIHIR+P GIDKIAARVQL
Sbjct: 6 IFIAFGVIVILAIVASTLYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGIDKIAARVQL 65

Query: 65 RLLQSEIIVETKTKDNVFTLNATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSV 124
RLLQSEIIVETKTKDNVFTLN+ATQYRVNE NVTDAYYKL+KPE+QIKSYIEDALRSSV
Sbjct: 66 RLLQSEIIVETKTKDNVFTLNATQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSV 125

Query: 125 PKLTLDLFEKKDEIALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 184
PKLTLDLFEKKDEIALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR
Sbjct: 126 PKLTLDLFEKKDEIALEVQHQAEEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 185

Query: 185 KRVAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLADSIQELKDANVTLT 244
KRVAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLA+SIQELK+AN++L
Sbjct: 186 KRVAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLN 245

Query: 245 EEQIMSILLTNQYLDTLNTFAINGNQTIFLPNNPEGVEDIRTQVLSALKTR 295
EEQIMSILLTNQYLDTLNTFA GNQT+FLPN P GVEDIRTQVLSALKT+
Sbjct: 246 EEQIMSILLTNQYLDTLNTFAAKGNQTLFLPNTPSGVEDIRTQVLSALKTK 296

SEQ ID 5810 (GBS231) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 7; MW 60.9kDa).

GBS231d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 5-7; MW 59kDa) and in Figure 239 (lane 11; MW 59kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 9; MW 34kDa) and in Figure 183 (lane 6; MW 34kDa). Purified GBS231d-GST is shown in Figure 246, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2115-

Example 1869

A DNA sequence (GBSx1977) was identified in *S. galactiae* <SEQ ID 5813> which encodes the amino acid sequence <SEQ ID 5814>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2305(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9291> which encodes amino acid sequence <SEQ ID 9292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13457 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
Identities = 259/514 (50%), Positives = 350/514 (67%), Gaps = 9/514 (1%)

Query: 1 MGMTMENGAKVEVSDKPAITTVGEVQILSKGVLMGARGNSGVITSQLEFRGFGQSIKDKKEEL 60
M ++M +GA+EV +G+VG LSKG+LMGARGNSGVI SQLFRGF ++I+ K+E+
Sbjct: 46 MNLSMTSGAREVEQMDTDDIGKVSALSGLLMGARGNSGVILSQLFRGFSGKNIETKKEI 105

Query: 61 TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
+ A A Q GV++AYKAVMKPVEGTILTV++ AA A+ AE+ D +M A + A
Sbjct: 106 NALFEAALQAGVDMAYKAVMKPVEGTILTVAKDAKKAMILAEKETDITALMTAVTEEA 165

Query: 121 KRALAKTPDMLPVLKEVGVDSSGQGLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
+ +L +TP++LPVLKEVGVDSSG+GL+ +YEGFL++L GE + KA ++ +MV+
Sbjct: 166 EASLNRTPELLPVLKEVGVDSSGKGLLCVYEGFLASLKGETVPQ---KAVLPSLDDMVS 222

Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYYEEFQGYLSNLGDSLLVND 240
AEHHK+ + TEDI++G+CTEVMV L Q +EF+ F+ LS GDSLLV+ D+
Sbjct: 223 AEHHKSAQSMNTEDIEFGFCTEVMVRLDQTK---REFDEGTFRQDLSQFGDSLLVIAD 279

Query: 241 EIVKVHVHTEDPGLVMQEGLYKGLVVKVENMRNQHDA---QMOKVEVEETVKETKEYG 297
+ KVH+H E+PG V+ YG L+K+K+ENMR QH + Q K ET + YG
Sbjct: 280 SLAKVHIHAEPEGVNLNYAQHYGELIKIKIENMREQHTSIISQESKPADNETPPAKQPYG 339

Query: 298 IIAVVGADGLAEIFKSQGVVDYIISGGQTMNPSTEDIVKAIEKVNARNVILPNNKNIFMA 357
I+ V G+G+A++FKS G +I GGQTMNPSTEDIV A++ VNA V ILPNN NI MA
Sbjct: 340 IVTVAMGEGIADLFKSIGASVVIIEGGQTMNPSTEDIVDAVKSVDNADTVFILPNNNSNIIMA 399

Query: 358 AQSAADVVDIPAADVETRTVPQGFTSLLAFDPKASLETNVADMTNSLSDVISGSVTLAVR 417
A AA VVD V+ +TVPGG ++LLAF+P + E N A+M +++ V SG VT +VR
Sbjct: 400 ANQASVVDQVVFIPAKTVPGQMSALLAFNPQEAANEANMLSAIQVKSGQVTFVSVR 459

Query: 418 DTTIDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDKMIDEDSEIVTIYVGEDGKQAL 477
DT IDG +I + D +G+++G I+ ++ + A K +MI ED EIVTI GED Q
Sbjct: 460 DTHIDGKDIKKGDFMGILNGTIIGTSENQLSAAKMLLSEMIGEDDEIVTILYGEDASQEE 519

Query: 478 AETLSEYLEETVEDVEVEIHQGDQPVYPYLSVE 511
AE L +L E YE++EVEIH G QP+Y Y++S E
Sbjct: 520 AEQLEAFLSEKYEEIEVEIHNGKQPLYSYIVSAE 553

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5635> which encodes the amino acid sequence <SEQ ID 5636>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1816(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2116-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 434/511 (84%), Positives = 475/511 (92%)

5
Query: 1 MGMTMENGAKEVSDKPAITVGEVQGILSKGVLMGARGNSGVITSQLFRGFGQSIKDKEEL 60
M MTM+NGAKEV+DKPA+TVGEVQG+LSKG+LMGARGNSGVITSQLFRGFGQSIK K+EL
Sbjct: 44 MSMTMDNGAKEVADKPAITVGEVQGMLSKGLLMGARGNSGVITSQLFRGFGQSIK GKDEL 103

10
Query: 61 TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
TG+DLA AFQ GVEVAYKAVMKPVEGTILTVSRGAATAALKKA+ TDDAVEVM+A L GA
Sbjct: 104 TGKDLAQAFQVGEVAYKAVMKPVEGTILTVSRGAATAALKKADLTDDAVEVMOAALDGA 163

15
Query: 121 KRALAKTPDMLPVLKEVGVVDSGGQGLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
K ALAKTPD+LPVLKEVGVVDSGGQGLVFIYEGFLSAL G+Y+ S DFKATPA M+EM+N
Sbjct: 164 KGALAKTPDLLPVLKEVGVVDSGGQGLVFIYEGFLSALNGDYVTSADFKATPANMSEMIN 223

20
Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYYEEFQGYLSNLGDSLLVND 240
AEHHK+VVGHVATEDI YGYCTE+MV LKQGPTYVKEFN+EFQGYLS LGDSLLVND
Sbjct: 224 AEHHKSVVGHVATEDITYGYCTEIMVALKQGPTYVKEFNDEFQGYLSGLGDSLLVND 283

25
Query: 241 EIVKVHVHTEDPGLVMQEGLYKGSLLVKVKNMRNQHDAQMQKVEVEETVKETKEYGIIA 300
EIVKVHVHTEDPGLVMQEGLYKGSLL+K+KV+NMRNQH+AQ+QK +VE+ E K++G+IA
Sbjct: 284 EIVKVHVHTEDPGLVMQEGLYKGSLLIKIKVDNMRNQHEAQVQKTDVEKNKAEVKDFGLIA 343

30
Query: 301 VVAGDGLAEIFKSQGVVDYIISGGQTMNPSTEDIVKAIEKVNARNVILPNNKNIFMAAQS 360
VVAG+GL+EIFK+QGVVDY+ISGGQTMNPSTEDIVKAIE VNA+ VIILPNNKNIFMAAQS
Sbjct: 344 VVAGEGLSEIFKAQGVVDYVISGGQTMNPSTEDIVKAIEAVNAKQVILPNNKNIFMAAQS 403

35
Query: 361 AADVVDIPAAVVETRTVPQGFTSLLAFDPAKSLETNVADMTNSLSDVISGSVTLAVRDTT 420
AA+VVDIPAAV TRTPVQGFTSLLAFDP+KSLE NVADM+ SLSDV+SGSVTLAVRDTT
Sbjct: 404 AAEVVDIPAAVVATRTVPQGFTSLLAFDPSKSLEDNVADMSTSLSDVVSGSVTLAVRDTT 463

40
Query: 421 IDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDKMIDEDSEIVTIYVGEDGKQALAE 480
IDGLEIHEND LGMVDGKI+VS PDME LK F+KMIDEDSEIVTI+VGE+G Q LAE
Sbjct: 464 IDGLEIHENDFLGMVDGKIIVSNPDMEATLKAAFEKMIDEDSEIVTIFVGEEDQDLAE 523

Query: 481 LSEYLEETIEDVEVEIHQGDQPVYPYLSVE 511
L+ YL ETYEDVEVEIHQGDQPVYPYLSVE
Sbjct: 524 LAGYLGETYEDVEVEIHQGDQPVYPYLSVE 554

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1870

45 A DNA sequence (GBSx1978) was identified in *S. agalactiae* <SEQ ID 5815> which encodes the amino acid sequence <SEQ ID 5816>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4771 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2117-

Example 1871

A DNA sequence (GBSx1979) was identified in *S.agalactiae* <SEQ ID 5817> which encodes the amino acid sequence <SEQ ID 5818>. This protein is predicted to be proliferating-cell nucleolar antigen P120. Analysis of this protein sequence reveals the following:

```

5   Possible site: 55
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3774(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9345> which encodes amino acid sequence <SEQ ID 9346> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC74905 GB:AE000278 putative nucleolar proteins [Escherichia
    coli K12]
    Identities = 87/229 (37%), Positives = 128/229 (54%), Gaps = 8/229 (3%)

20 Query: 63  GKSIEHTTGLVYSQEPAAQ--IVAQIAEPQEGMKVLDLAAAPGGKTTHLLSYLNNTGLLV 120
    G + EH +GL Y QE ++ + A A+ +V+D+AAAPG KIT + + +NN G ++
    Sbjct: 89  GSTAEHLISGLFYIQEASSMLPVAALFADGNAPQVRMDVAAAPGSKTTQISARMNNEGAIL 148

    Query: 121 SNEISNKRKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPCSGEGMFRKD 180
    +NE S R K+L N+ R G NV +T+ + FD I+ D PCSGEG+ RKD
    Sbjct: 149 ANEFSASRVKVLHANISRCGISNVALTHFDGRVFGAAVPEMFDAILLDAPCSGEGVVRKD 208

    Query: 181 PQAIFYWHKDYPTCAQLORDILKEAIKMLAHGGILVYSTCTWSPEENEVNVNLLQEY- 239
    P A++ W + E A QR+++ A L GG LVYSTCT + EENE V WL + Y
    Sbjct: 209 PDALKNWSPESNQEIAATQRELIDSAFHALRPGGTLVYSTCTLNQEENEAVCLWLKETYP 268

    Query: 240 ---DYLELVDIPKLNMGVEGINVPQVARMYPHHFQEGGQFVAKLRDTRS 285
    ++L L D+ G + + ++P + EG FVA+LR T++
    Sbjct: 269 DAVEFLPLGDL--FPGANKALTEGFLHVFPQIYDCEGFFVARLRKTQA 315
35

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5819> which encodes the amino acid sequence <SEQ ID 5820>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2316(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 213/311 (68%), Positives = 254/311 (81%), Gaps = 3/311 (0%)

50 Query: 1  MKLPNEFIEKYQTILKDEAEAFFDSFEQKPI SAYRTNPLKEKQLDFPNAIPSTPWGHYK 60
    M LP EFI YQ IL E E F SF Q+P++A+R NPLK + F + IP+T WG+YGK
    Sbjct: 2  MSLPKFEFINTYQAILGKELEDFLASFNQEPVNAFRINPLKNQLKTFEHPIPNLTWGYYGK 61

    Query: 61  ISGKSIEHTTGLVYSQEPAAQIVAQIAEPQEGMKVLDLAAAPGGKTTHLLSYLNNTGLLV 120
    +SGKS EH +GLVYSQEPAAQ+VAQ+A PQ+G +VLDLAAAPGGK+THLL+YL+NTGLLV
    Sbjct: 62  LSGKSPEHVSGLVYSQEPAAQMQVAAPQKGSRVLDLAAAPGGKSTHLLAYLDNTGLLV 121

    Query: 121 SNEISNKRKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPCSGEGMFRKD 180
    SNEIS KRK+LVEN+ERFGARNV+VTNES+ RLAK F+ +FD IVFDGPCSGEGMFRKD
    Sbjct: 122 SNEISKRSKVLVENIERFGARNVVVTNESADRLAKVFSHYFDITIVFDGPCSGEGMFRKD 181

```

-2118-

Query: 181 PQAIQYWHKDYPTCAQLQORDILKEAIKMLAHGGILVYSTCTWSPEENEVNVNLLQEQD 240
 P AIQYWH YP ECA+LQ+ IL++A+ ML GG L+YSTCTW+PEENE+VV WLL+ Y
 Sbjct: 182 PDAIQYWHHGYPAECAKLQKSILEDALAMLKPGGELIYSTCTWAPENEDVVQWLLLETYT 241

Query: 241 YLELVDIPKLNMGVGINVPQARMYPHFFQEGQFVAKLRDTRSKAQKIKPKAQKIN- 299
 +LELVD+PKLNGMV GI +P+ ARMYPH +QEGQFVAKL+D R +E Q K KA K N
 Sbjct: 242 FLELVDVPKLNMGVSGIGLPETARMYPHRYQEGQFVAKLKDKR-QEGQSTKLKAPKSNL 300

Query: 300 -KMQLQLWQQF 309
 K QL+LW+ F
 Sbjct: 301 IKDQLRLWKMF 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1872

A DNA sequence (GBSx1980) was identified in *S.agalactiae* <SEQ ID 5821> which encodes the amino acid sequence <SEQ ID 5822>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4111(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24940 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 86/240 (35%), Positives = 133/240 (54%), Gaps = 10/240 (4%)

Query: 6 DFAKQLVYKAGQFIKSEMONTFDVEEKSFRDDLVTSLDKKTQKLLIQEIIQHYPDDNILA 65
 + AK+ + +AG I M + +E KS +DLVT++DK+T+K I I + +P IL
 Sbjct: 9 EIAKKWIREAGARITQSMHESLTIEFKSNPNDLVTNIDKETEKFFIDRIQETFFPGHRILG 68

Query: 66 EE---DBVRSPIAQGNVWVLDPIDGTNVNFVQKDNFAVMLAYYEEGVGFIIYDVMADI 122
 EE D + S +G VW++DPIDGT+NF+ Q+ NFA+ + +E G G+ G+IYDV+ D
 Sbjct: 69 EBGQGDKIHS--LEGVWIIDPIDGTNNFVHQQRNFAISIGIFENGEGKIGLIYDVVHDE 126

Query: 123 LYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAEN---DCGIAHLASETLGVRI 178
 LY Y N+ K+ P +E +E +L +N+ EN +A L G R
 Sbjct: 127 LYHAFSGRGAYMNETKLAPLKETVIEEAILAINATWVTENRRIDQSVLAPLVKVRGTR 186

Query: 179 YGGAGISMAKVMQKLLAYFSY-IQPWDYAAKIMGETLGFLLTLDGEEPNYSTRQKVM 237
 YG A + +A V G++ AY + + PWDYAA ++ +G T T++GE + V+
 Sbjct: 187 YGSAALELANVAAGRIDAYITMRLAPWDYAGCVLLNEVGTTYTTIEGEPFTFLENHSLV 246

A related GBS nucleic acid sequence <SEQ ID 10937> which encodes amino acid sequence <SEQ ID 10938> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5823> which encodes the amino acid sequence <SEQ ID 5824>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2119-

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/253 (61%), Positives = 205/253 (80%)

```

5   Query: 1  MDAKFDFAKQLVYKAGQFIKSEMQNTFDVEEKSFRDDLVTSLDKKTQKLLIQEIQHYPD 60
      ++ K+ FA+Q++ +AG FIKS+M    D++ K++FDDLVT++D++TQ+LL+  I Q YP
      Sbjct: 8  LETKYAFARQIIKEAGLFIKSKMSEQLDIQVKTQFDDLVTNVDQETQQLLMDRIHQITYPC 67

10  Query: 61  DNILAEEDBVRSPIAQGNVWVLDPIDGTNVNFIVQKDNFAVMLAYYEEGVGQFGIYDVMA 120
      D ILAEE++VR PI QGNVWV+DPIDGTNVNFIVQ  FAVM+AYYE+G+GQFG+IYDVMA
      Sbjct: 68  DAILAEENDVRHPINQGNVWVIDPIDGTNVNFIVQGSQFAVMIAYYEQSIGQFGLIYDVMA 127

15  Query: 121 DILYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAENDCGIAHLASETLGVRIYG 180
      D L +GGG F+V N K+ +QE PLER L+G N+ M+A ND +AHL ++TLGVR+YG
      Sbjct: 128 DQLLAGGGDFEVTLNGDKLPAYQEKPLERSLIGCNAGMFARNDRNLAHLIAKTLGVRVYG 187

      Query: 181 GAGISMAKVMQKLLAYFSYIQPWDYAAAKIMGETLGFLLTLDGEEPNYSTRQKVMFLP 240
      GAGI M KVM+ +LLAYFS+IQPWDYAAAK++G+ LG+ LLT+DG EP++ TRQK+MF+P
      Sbjct: 188 GAGICMVKVMKQELLAYFSFIQPWDYAAAKVLGDKLGYVLLTIDGYEPDFQTRQKIMFVP 247

20  Query: 241 KSKLNLIQSYLTK 253
      K +L I S+LTK
      Sbjct: 248 KCQLTRIASFLTK 260

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1873

A DNA sequence (GBSx1981) was identified in *S.agalactiae* <SEQ ID 5825> which encodes the amino acid sequence <SEQ ID 5826>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 16
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4131(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:AAC24938.GB:AF012285 unknown [Bacillus subtilis]
    Identities = 33/78 (42%), Positives = 50/78 (63%)

      Query: 13 YSYPLDPSWNTEDITKVLRFNLQVEHAYENSIKVDDLDSYKEFKKVVSKAQEKQIDRE 72
      Y YP++ W TE+ V+ F QVE AYE ++LL +Y+ FK++V KA+EK++ E
      Sbjct: 3  YQYPMNEDWTTEEAVDVIAFFQQVELAYEKGADREELLKAYRRFKEIVPGKAEEKKLCGE 62

45  Query: 73 FQRTSGYSTYQAVKAAQQ 90
      F+ S YS Y+ VK A++
      Sbjct: 63 FEEQSTYSPYRTVKQARE 80

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5827> which encodes the amino acid sequence <SEQ ID 5828>. Analysis of this protein sequence reveals the following:

```

55  Possible site: 60
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-2120-

Identities = 59/91 (64%), Positives = 70/91 (76%)

Query: 9 ISSNYSYPLDPSWNTEDITKVLRLNQEVEHAYENSIKVDDLLDSYKEFKKVVKSQAQEKQ 68
 +S NY YPLD SW+TE+I+ VL FLN+VE AYE + LLDSYK +K +VKSKAQEKQ
 Sbjct: 5 MSGNYYYPLDLWSSTEEISSVLHFLNKVELAYEKKVDAKQLLDSYKTYKTTVKSKAQEKQ 64

Query: 69 IDREFQRTSGYSTYQAVKAAQQQAKGFISLG 99
 IDR+FQ+ SGYSTYQ VK A+ KGF SLG
 Sbjct: 65 IDRDFQKVSGYSTYQVVKAKAKAIEKGFFSLG 95

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1874

A DNA sequence (GBSx1982) was identified in *S.agalactiae* <SEQ ID 5829> which encodes the amino acid sequence <SEQ ID 5830>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence (or aa 1-18)

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0952(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF21893 GB:AF103794 unknown [Listeria monocytogenes]
 Identities = 74/126 (58%), Positives = 101/126 (79%)

Query: 1 MITLFLSPSCTSCRKARAWLSKHEVAFEHNIITSPLNKEELLQILSFTENGTEIDIISTR 60
 M+TL+ SPSCCTSCRK+RAWL +H++ ++E NI + PL+ +E+ +IL TE+GT++IISTR
 Sbjct: 1 MVTLYTSPSCTSCRKSRAWLEEHDIYPKERNIFSEPLSLDEIKEILRMTEGTDEIISTR 60

Query: 61 SKVFQKLAIIDVDELSTSSLMELISENPSLLRRPIILDKKRMQIGFNEDEIRAFLPRDYRK 120
 SK FQKL +D+D L L ELI +NP LLRRPII+D+KR+Q+G+NEDEIR FLPR R
 Sbjct: 61 SKTFQKLNVDLSLPLQQLFELIQKNPGLRRPIIIDEKRLQVGYNEDEIRRFLPRRVRT 120

Query: 121 QELKQA 126
 +L++A
 Sbjct: 121 YQLREA 126

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5831> which encodes the amino acid sequence <SEQ ID 5832>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0511(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 112/134 (83%), Positives = 127/134 (94%)

Query: 1 MITLFLSPSCTSCRKARAWLSKHEVAFEHNIITSPLNKEELLQILSFTENGTEIDIISTR 60
 M+TLFLSPSCTSCRKARAWL KHEV F+EHNIITSPL+++EL+ ILSFTENGTEIDIISTR
 Sbjct: 1 MVTFLFLSPSCTSCRKARAWLVKHEVDFOEHNIITSPLSRDELMSILSFTENGTEIDIISTR 60

Query: 61 SKVFQKLAIIDVDELSTSSLMELISENPSLLRRPIILDKKRMQIGFNEDEIRAFLPRDYRK 120
 SKVFQKL IDV+ELS S L++LI++NPSLLRRPII+D+KRMQIGFNEDEIRAFL R DYRK
 Sbjct: 61 SKVFQKLDIDVEELSISDLIDLIAKNPSLLRRPIIMDQKRMQIGFNEDEIRAFLSRDYRK 120

-2121-

Query: 121 QELKQATIRAEIEG 134
 QEL+QATI+AEIEG
 Sbjct: 121 QELRQATIKAEIEG 134

5

SEQ ID 5830 (GBS232) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 10; MW 16.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 2; MW 42kDa).

GBS232-GST was purified as shown in Figure 207, lane 7.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1875

A DNA sequence (GBSx1983) was identified in *S.agalactiae* <SEQ ID 5833> which encodes the amino acid sequence <SEQ ID 5834>. Analysis of this protein sequence reveals the following:

15 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5835> which encodes the amino acid sequence <SEQ ID 5836>. Analysis of this protein sequence reveals the following:

25 Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1768(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 210/308 (68%), Positives = 252/308 (81%)

35 Query: 1 MKIHYINDYKDIQAKEDCVLVLGYFDGLHLGHKALFDKAKKIATEKNLKIVVLTFNETPR 60
 M+I YI DY+DI ++D VL+LGYFDGLH GHKALFDKA+++A ++ LK+VV TF E+P+
 Sbjct: 1 MEIEYIKDYRDINQEDDTVLILGYFDGLHRGHKALFDKAREVANKEGLKVVVFTFTESPK 60

40 Query: 61 LTFARFQPELLLLHLTSPEKRSEKFOEYGVDELYLMNFTSHFSKVSSDLFIKKYIYGLRAK 120
 L F+RF PELL LH+T P+KR EKF +YGV++LYL++FTS FSKVSSD FI YI L+AK
 Sbjct: 61 LAFSRFSPPELLLHITYPKRYEKFADYGVNKLVLVDFTSKFSKVSSDHFITHYIKNLKAK 120

45 Query: 121 AAVVGFDYKFGHNRTSGDYLRNFKGPVYIIDEISEGGEKISSTRIRQLITEGNVEKANQ 180
 VVGFDYKFGHNRT DYL RNF+G VY I+EI E KIS+T IR+LI EGNV KAN
 Sbjct: 121 HIVVGFDYKFGHNRTSDSYLTRNPEGQVYITIEIKEDHRKISATWIRKLIQEGNVVKANH 180

50 Query: 181 LLGYEFSTCGMVVHGDARGRTIGFPTANLAPINRTYLPADGVYISNVLINGKYYRAMTSI 240
 LLGY+ ST G VVHGDARGRTIGFPTANLAPI+ TYLPADGVY++NV++ K YR+MTS+
 Sbjct: 181 LLGYDLSTRGRVVHGDARGRTIGFPTANLAPIDNTYLPADGVYVTNVIVANKIYRSMTSL 240

55 Query: 241 GKNITFGGTTELRLLEANIFDFDGDYGETIEIFWLKRIEMVKFNGIDDLVKQLKKDKKEIA 300
 GKN+TFGG ELRL NIFDFD +IYGE IEI WL +IR+M KF GI+DL +L+ DK A
 Sbjct: 241 GKNVTFGGKELRLVNIIFDFDEEYGEIIEIVWLDKIRDMKFEIGIEDLTDREYDKRTA 300

Query: 301 LNWKDSQ 308

-2122-

LNWKKDS+
Sbjct: 301 LNWKKDSK 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1876

A DNA sequence (GBSx1984) was identified in *S.agalactiae* <SEQ ID 5837> which encodes the amino acid sequence <SEQ ID 5838>. This protein is predicted to be tRNA pseudouridine 5S synthase (truB). Analysis of this protein sequence reveals the following:

10 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
15 bacterial cytoplasm --- Certainty=0.2576 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9817> which encodes amino acid sequence <SEQ ID 9818> was also identified.

20 The protein has homology with the following sequences in the GENPEPT database.

 >GP:BAB06129 GB:AP001515 tRNA pseudouridine 5S synthase [Bacillus halodurans]
 Identities = 145/283 (51%), Positives = 191/283 (67%), Gaps = 12/283 (4%)

25 Query: 2 ITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYMT 61
 +TGI+ L K GMTSHD V KLR++L TTK+GH GTLDPDV GVLP+ +G AT+V +YM+
 Sbjct: 3 MTGILPLAKPRGMTSHDCVAKLRRLKTKKVGHTGTLDPDVVGVLPCIGHATKVAQYMS 62

30 Query: 62 ESGKIYEGETLGYATSTEDSSGEVISRTPLTQSDLSEDDVVDHAMKSFTGPITQVPPMYS 121
 + K YEGE+T+G++T+TED SG+ + T Q E VVD + +F G I Q+PPMYS
 Sbjct: 63 DYPKAYEGEVTVGFSSTTTEDRSQDIVE-TKTIQQPFVEAVDQVLATFVGEIKQIPPMYS 121

35 Query: 122 AVKVNGKKLYEYARSGEEVERPKRQITISEFRRTSPLYFEKGICRFSFYVSCSKGTYVRT 181
 AVKV GK+LYEYAR+G VERP+R +TI R S + +E+G+CRF F VSCSKGTYVRT
 Sbjct: 122 AVKVRGKRLYEYARAGITVERPERTVTIFSLERMSDIVYEEGVCRFRFNVSCSKGTYVRT 181

40 Query: 182 LAVDLGIKLG YASHMSFLKRTSSAGLSITQSLTLEEINEKYKQ-EDFSFLLPIEYGVLDL 240
 LAVD+G LGY +HMS L RT S S+ + T E+ E+ +Q E S LLPIE +LD+
 Sbjct: 182 LAVDIGKALGYPAHMSDLVRTKSGPFSLEECFTFTELEERLEQEGSSLLLPITAILDI 241

40 Query: 241 PKVNLTEEDKVEISYGR-----RILLENEADTLAAFYE 273
 P+V + +E + +I +G R + NE L A Y+
 Sbjct: 242 PRVQVNKEIEEKIRHGAVLPQKWFNHPRTVYNEEGALLAIYK 284

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5839> which encodes the amino acid sequence <SEQ ID 5840>. Analysis of this protein sequence reveals the following:

 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2698 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

55 Identities = 201/295 (68%), Positives = 246/295 (83%), Gaps = 2/295 (0%)

-2123-

Query: 1 MITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60
 MI GIINLKKEAGMTSHDAVFKLRK+L KKIGHGGTLDPDVVGVLPIAVGKATRVIEYM
 Sbjct: 1 MINGIINLKKEAGMTSHDAVFKLRKLLQEKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60

5 Query: 61 TESGKIYEGETLGYATSTEDSSGEVISRTPLTQSDLSEDVVDHAMKSFTGPITQVPPMY 120
 TE+GK+YEG++TLGY+T+TED+SGEV++R+ L + L+E++VD M +F G ITQ PPMY
 Sbjct: 61 TEAGKVYEGQVTLGYSTTTEDASGEVVARSSL-PAVLTEELVDQTMITFLGKITQTPPMY 119

10 Query: 121 SAVKVNKGKLYEYARSGEEVERPKRQITISEFRRTSPLYF-EKGICRFSFYVSCSKGTYY 179
 SAVKVNKG+KLYEYAR+GE VERP+R++TIS F RTSPL F E G+CRFSF V+CSKGTYY
 Sbjct: 120 SAVKVNKGKLYEYARAGESVERPRREVTISLFFERTSPLNFTEDGLCRFSFKVACSKGTYY 179

15 Query: 180 RTLAVDLGIKLYASHMSFLKRTSSAGLSITQSLTLEEINEKYKQEDFSFLLPIEYGVLD 239
 RTLAVDLG LG SHMSFL+R++SAGL++ + TL EI + +++ SFLPIEYGV D
 Sbjct: 180 RTLAVDLGRALGVESHMSFLQRSASAGLTLETAYTLGEIADMVSKQEMSFLPIEYGVAD 239

20 Query: 240 LPKVNLTEDKVEISYGRILLNEADTLAIFYENRVIAILEKRGNEFKPHKVLL 294
 LPK+ + + + EIS+GRR+ L ++ LAAF+ +VIAILEKR E+KP KVL+
 Sbjct: 240 LPKMVIDDTLTELISFGRRSLPSQEPLLAAPHGEKVIAILEKRDQYKPKKVL 294

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1877

A DNA sequence (GBSx1985) was identified in *S.agalactiae* <SEQ ID 5841> which encodes the amino acid sequence <SEQ ID 5842>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2776(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9819> which encodes amino acid sequence <SEQ ID 9820> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12871 GB:Z99109 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 39/145 (26%), Positives = 68/145 (46%), Gaps = 7/145 (4%)

40 Query: 3 MKIRTATLDDSEKLVPLYQELG---YAISSLSEIQSILKVILTHSDYGLIAEDNGKLLA 58
 M IR A D+ + PL+ + A L ++ LK L + + LIAE+NG+ +
 Sbjct: 1 MNIRQAKTSDAAAIAPLFNQYREFYRQASDLQGAFLKARLENHESVILIAEENGFIG 60

45 Query: 59 FVGYHKLYFFEKSGTYRILALVVNEKHRRKGIASQLINHVQLAKTDGSEVLALNSSLK 118
 F + + Y + L V R KG +L++ K A +G++ L L + +
 Sbjct: 61 FTQLYPTFSSVSMKRIYILNDLFVVPHARTKGAGGRLLSAAKDYAGQNGAKCLTQT--E 118

Query: 119 EYRQEAYHFYENLGFKKVSTGFSYY 143

+ ++A YE G+++ TGF +Y

50 Sbjct: 119 HHNRKARSLEYQNGYEE-DTGFVHY 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5843> which encodes the amino acid sequence <SEQ ID 5844>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0962(Affirmative) < succ>

-2124-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 37/126 (29%), Positives = 64/126 (50%), Gaps = 16/126 (12%)

Query: 18 PLYQE-----LGYAISLSEIQSILKVILTHSDYGFLLIA--EDNGKLLAFVG--YHKLYF 67
 P+ QE LGY +SL ++ + ++ + FL +D +LL +V Y LY
 10 Sbjct: 11 PMLQEINAKALGYLVSLDLLERQYERLIEDCHHYFLAYADKDTNQLLGYVHAERYETLY- 69

Query: 68 FEKSGTYTYRILALVVNEKHRRKGIASQLINHVQKLAKTGDGSEVLAINSSLKEYRQEAYHF 127
 + +L L V ++R+GI S L+ ++ A+ +G + LNS+ +R+EA+ F
 15 Sbjct: 70 ---ASDGLNLLGLAVLPAYQRRGIGSALLRALESQARQEGIAFIRLNSA--SHRKEAHAF 124

15 Query: 128 YENLGF 133
 Y NL +
 Sbjct: 125 YRNLDY 130

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 20 vaccines or diagnostics.

Example 1878

A DNA sequence (GBSx1986) was identified in *S.agalactiae* <SEQ ID 5845> which encodes the amino
 acid sequence <SEQ ID 5846>. Analysis of this protein sequence reveals the following:

25 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1659(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 28-30

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAF30776 GB:AE002133 conserved hypothetical [Ureaplasma
 urealyticum]
 Identities = 106/440 (24%), Positives = 206/440 (46%), Gaps = 65/440 (14%)

40 Query: 13 FAINESEYHQLLEQIRGDAFDKEVSEERLEKERLILGEQAKNQLQEVVVE-KDKEIAKLQY 71
 F N+ +Y++L++Q +D LEK+R L E+ KN+ + + KD + K
 Sbjct: 71 FLANDRDYNELVKQ----RYD-----LEKQDELKEKLKNEGKATAHFKDSDEYKNLI 120

45 Query: 72 KVKQFLIEKDNLLKDNFYQLAEQLNQKDMMLRD-----LENQIDRLRLHEHENSILQEA 123
 K ++ + + ++ NE +++ ++ L+ L+N I + ++ +N+ + A
 Sbjct: 121 KAQEKINSLNKTTIESNEQSYKKEIENIELKLKSQFDEETKSLKNTIAKQEIKLDNAEKMA 180

50 Query: 124 LTKVERE-----RDAIQNLHIQ-----EKEKDLALASVKS DY 156
 + + +D I + I+ E +K + + ++S
 Sbjct: 181 IINFKESNEYQKI IKDKIDLDIEIEKLKFAIQAHEDNMKA AKENWESKKIVEIKELESKK 240

55 Query: 157 EVQLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKRHLAAPPNAYFEKDNTLSSR 216
 + ++ E +E K K+ + K VGE LE + + +F++ + P+ F K N
 Sbjct: 241 DKEIHKLTESIEQLKREKSS-NVKLVGEELEQWLKNKFDETFYSFSCPDMTFTKINEAID- 298

60 Query: 217 GSKGDFIY-----REKDENDLEFL-SIMFEMKNESDDTIKKHKNE DFFKELDKDRREKS 269
 G K DF+ +E +D + + S E K E D K KN +K+LD+DR +
 Sbjct: 299 GKADFLLEFFDFGKEMSNDKKLIFSATIEAKTEFFDNQKGTKNSAHYKKLDQDRINQK 358

60 Query: 270 CEYAVLVITMLEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNLTLYKQEL 329
 EYA+LVT LE ++ + ++ ++Y M+ +RPQ+FI L+ ++RN A TLK K
 Sbjct: 359 SEYAILVTELEPEDHF---VIKKINEYKNMFAVRPQYFIPLVDMIRNFA--TLKAKINS 412

-2125-

Query: 330 ALMKEQNIDITHFEEDLDIFKNAFAKN-YNSASKNFQKAIDEIDKSIKRMEAV-KAALTT 387
 +++ + D .EE+LD K N + +K ID+ IK+ E++ ++A
 Sbjct: 413 QIIRYE--DRAKIEENLDELKKDIVDNTLKYINDKTKKIIDSKAIIKKAESIEESAEDI 470

Query: 388 SENQLRLANNKLDDVSVKKL 407
 +L K++++++K+
 Sbjct: 471 INKKLNTLKKKINELTIRKI 490

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5847> which encodes the amino acid sequence <SEQ ID 5848>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3192(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 310/445 (69%), Positives = 352/445 (78%), Gaps = 22/445 (4%)

- Query: 1 MNEIKCPHCGTAFAINSEYHQLLEQIRGDAFDKEVSEERLEKERLILGEQAKNQLQEVVV 60
 MNEIKCPHC T F INESEY QLLEQ+RG AFD+E+ +RL E +L E+AK+QL EVV
 Sbjct: 1 MNEIKCPHCHTLFTINESEYSQLEQVRGQAFDEELKKRLINETALLEEKAKHQLHEVVA 60
- Query: 61 EKDKETIAKLQYKVKQF-----LIEKDNLL-----KDNEYQLAEQLNQK 98
 +K+ I L +++Q L +KD L+ N +LA QL +K
 Sbjct: 61 KKETAITSLTINQLEQIEKEQAYLRQEELAKKDQLIASLEAKLKLASQNALELANQLAEK 120
- Query: 99 DMMLRDLENQIDRLRLEHENSQLEALTKVERERDAIQNLHIQEKEKDLALASVKSDYEV 158
 D + L NQ+D+L LE + + Q L +E+ERD I+NQL +Q KE +L+LASV+SDYE
 Sbjct: 121 DKEVVSILTQDLKALEKDQATFQSKLATIEKERDGIKNQLALQAKESELSLASVRSYEA 180
- Query: 159 QLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKVRHLAFPNA YFEKDNTLSSRGS 218
 QLKAANEQVEFYKNFKAQQSTKA+GESLE YAETEFNKVR AFPNA F KDN LSSRGS
 Sbjct: 181 QLKAANEQVEFYKNFKAQQSTKAIGESLELYAETEFNKVRSYAFPNASFVKDNQLSSRGS 240
- Query: 219 KGDFIYREKDNEDLEFLSIMFEMKNESDDTIKKHKNE DFFKELDKDRREKSCEYAVLVTM 278
 KGD+IYRE D N +E LSIMFEMKNE+D T KHKN DFFKELDKDRREK CEYAVLV+M
 Sbjct: 241 KGDYIYREVDANGVEILSIMFEMKNEADTTKTKHKNSDFFKELDKDRREKDCYAVLVSM 300
- Query: 279 LEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYKQELALMKEQNID 338
 LEADNDYYNTGIVDVSH+Y KMYV+RPQ FIQLIGILRNAALN+L YKQELAL+KEQNID
 Sbjct: 301 LEADNDYYNTGIVDVSHYQKMYVVRPQLFIQLIGILRNAALNSLHYKQELALVKEQNID 360
- Query: 339 ITHFEEDLDIFKNAFAKNYNSASKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNK 398
 ITHFEEDLD FKNAFAKNY SAS NF+KAIDEIDKSIKRME VK LTTSENQLRLANNK
 Sbjct: 361 ITHFEEDLDQFKNAFAKNYQSASNNFKKAIDEIDKSIKRMEEVKRFLLTTSENQLRLANNK 420
- Query: 399 LDDVSVKKLTRKNPTMKAKFDALKD 423
 L+DVSVKKLTR+NPTM+ KF+ALKD
 Sbjct: 421 LEDVSVKKLTRQNPTMREKFEALKD 445

- 55 SEQ ID 5846 (GBS304) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 206, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2126-

Example 1879

A DNA sequence (GBSx1987) was identified in *S.agalactiae* <SEQ ID 5849> which encodes the amino acid sequence <SEQ ID 5850>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1845(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5851> which encodes the amino acid sequence <SEQ ID 5852>. Analysis of this protein sequence reveals the following:

```

15   Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.2492(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25   Identities = 113/180 (62%), Positives = 141/180 (77%)

Query: 16  LSELVDCFKGKAVPSKAEAGDIRIINLSDMSPLGIDYHNLRTFQDEQRSLLKYLLQEGDV 75
          L +VDCFKGKAV SK  GD+ +INLSDM LGI YH LRFTQ ++R LL+YLL++GDV
Sbjct: 18  LGTVVDCFKGKAVSSKVVPDVGILINLSDMGTLGIQYHQLRTFQMDRRQLLRILLEGDV 77

30   Query: 76  LIASKGTVKKVAIFEEQDYPVVASANITILRPTQHIRGYLKLFFDSEEGQQALENANKG 135
          LIASKGT+KKV +F +Q+ VVAS+NIT+LRP + +RGYY+K F DS GQ L+ A+ G
Sbjct: 78  LIASKGTLKKVCVFHKQNRDVASSNITVLRPQKLLRGYYIKFFLDSPIGQALLDVADHG 137

35   Query: 136 KAVMNISTKELLNIAIPSIPLFRQDYLIQRYKQGLNDYKRKIARAEQEWERIQNDIRQQL 195
          K V+N+STKELL+I IP IPL +QDYLI Y +GL DY RK+ RAEQEWE IQN+I++ L
Sbjct: 138 KDVINLSTKELLDIPIPIPLVKQDYLINHYLRGLTDYHRKLNRAEQEWYIQNEIQKGL 197

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1880

A DNA sequence (GBSx1988) was identified in *S.agalactiae* <SEQ ID 5853> which encodes the amino acid sequence <SEQ ID 5854>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 15
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.43    Transmembrane    62 - 78 ( 55 - 82)
      INTEGRAL    Likelihood = -2.87    Transmembrane   130 - 146 ( 130 - 150)
      INTEGRAL    Likelihood = -1.28    Transmembrane    37 - 53 ( 37 - 53)

50   ----- Final Results -----
      bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2127-

A related GBS nucleic acid sequence <SEQ ID 9347> which encodes amino acid sequence <SEQ ID 9348> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5      >GP:CAA22372 GB:AL034446 putative transmembrane protein
      [Streptomyces coelicolor A3(2)]
      Identities = 38/139 (27%), Positives = 64/139 (45%), Gaps = 5/139 (3%)

Query: 15  SASVEILCRGWLLPVSATKYSKIVSVSISSIFFGLLHSANNHVSLSIFNLCL-FGLFLS 73
      +A+ E++ RG L +      +++ ++ + FGL+H N +L + + G L+
10  Sbjct: 143 AATEEVVFRGVLFRIIEEHIGTYLALGLTGLVFGMLHLLNEDATLWGALAIATIEAGFMLA 202

Query: 74  LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSGEPMLSNSLVHVKTYGADWISGGKFGVE 133
      N+W G+H WN G VF VSG S L+ G ++CG FG E
15  Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVSNGND-SEGLLDATMSGPKLLTGDFGPE 261

Query: 134 GSMIT---SIVLIVACYWL 149
      GS+ + ++L + WL
20  Sbjct: 262 GSVYSVGFVGLLTLVFLWL 280

```

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1881

25 A DNA sequence (GBSx1989) was identified in *S.agalactiae* <SEQ ID 5855> which encodes the amino acid sequence <SEQ ID 5856>, which is a methylase gene homolog. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

30  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2192(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35  RGD motif: 264-266

```

A related GBS nucleic acid sequence <SEQ ID 9929> which encodes amino acid sequence <SEQ ID 9930> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:BAA87672 GB:AB016260 Hypothetical gene, methylase gene homolog
      [Agrobacterium tumefaciens]
      Identities = 358/1238 (28%), Positives = 595/1238 (47%), Gaps = 99/1238 (7%)

45  Query: 1072 KEVARIKGMVDIRNAYQEVIAIQRYDYDKETFNHLLGKLNRTYDSFVKHYGYLNSAV-- 1129
      K V I+ ++ IR+A +EV+ Q      + L +L + SFV+ +G +N
      Sbjct: 497 KHVRIIRKLIPIRDVREVLKAQEL----DRPWKDLQVRLRVAVSSSFVRDFGPINHTTVS 552

Query: 1130 -----NRNLFSDDDKYSLLASLEDESL--DPSGKSVIYTKSLAFEKAL 1170
      N F D L+AS+ED L D + I+T E+ +
50  Sbjct: 553 ITEDPESGETRESHRRPNLQPFADDPDCWLVASIEDYDLENDTAKPGAIFT-----ERVI 607

Query: 1171 VRPEKEVKKVHTALDALNSSLADGRGVDFAYMMSIQVESQMTLIEELGDLIMPDPPEKYL 1230
      P V + +A DAL L+ VD ++ + + ++ ELG I DP
55  Sbjct: 608 SPPAPPV--ITSAADALAVVLNERGRVDLDHIAELLHRDPD-DVVAELGSAIFRDP---- 660

```

-2128-

- Query: 1231 NGELTYVSRQDFLSGDVVTKLEVVDLFVKQDNQDFNWSHYAGLLEAIKPARITLADIDYR 1290
+ ++ +LSG V KL+V + D ++ L ++P + +DI R
Sbjct: 661 -ADGSWQMADAYLSGPVRDKLKVAAAAALDPV---YNNRVTALAGVQPVDLRPSDITAR 716
- 5
Query: 1291 IGSRWIPLAVYKFAQETFMGKAYELSDQ-EVATVLEVSPIDGVITYQSKFAYTYSNATD 1349
+G+ WIP A F +E MG + E+A+ + G + A T TD
Sbjct: 717 LGAPWIPAADVVAFVKE-MMGTDIRIHMPELASWTVEARQLGYLA-----AGTSEWGTD 770
- 10
Query: 1350 RSLGVPASRYDSGRKIFENLLNSNQPTTTKQVVEGDKKNVTDVEKTTVLRAKETHLOEL 1409
R ++ + LNS P I + +GD ++ V +V T + K +++
Sbjct: 771 RR-----HAGELLSDALNSRVPQIFDTIRDGDSERRVLNVVDTEAAKEKLHKIKDA 821
- 15
Query: 1410 FQGFVAKYPEVQQMIEDTYNRLYNRTVSKSYDGSHLTIDGLAQNISLRPHQKNAIQRIVE 1469
FQ ++ P+ + YN +N + + G HL + G + L HQK I RI+
Sbjct: 822 FQRWIWSDPDRTDRLARVYNDRFNNIAPRKFSGDHLNLPASGAFVLYGHQKRGWIWRIIS 881
- 20
Query: 1470 EKRALLAHEVSGSKTTLTMLGAGFKLKEGLMVHKPLYVVPSSLTQAQFQEIIMKFFPTKKVY 1529
LAH VG+GKT+TM + + + LG++ K + VVP AQ +E + +PT ++
Sbjct: 882 SGSTYLAHAVGAGKTMITMAASIMEQRRIGLIAKAMQVPGHCLAQAAREFLALYPTARIL 941
- 25
Query: 1530 VITKKDFAKAKRKQFVSRITIGDYDAIVIGDSQFEKIPMSREKQVITYINDKLEQLREIKL 1589
V + +F+K KR +F+SR T +DAI+I S F I + + I+D+LE + L
Sbjct: 942 VADETNSKDKRARFSLRAATATWDIIITHSAFRFIGVPAAFESQMIHDELELYETLLL 1001
- 30
Query: 1590 GSDSDYTV--KEAERSIKGLEHQLEELQKLERDTFIEFENLGIDFLFVDEAHHFKNIRPI 1647
+ + V K ER +GL+ +LE L +D + +G+D + VDEA F+ +
Sbjct: 1002 KVEDEDRVSRKRLERLKEGLQERLEALST-RKDDLLTIAEIGVDQIIVDEAQEFRLKSFA 1060
- 35
Query: 1648 TGLGNVAGITNTTSKKNVDMEMKVRQVQAEHGDNRNVFATGTPVSNISSELTMMDYIQP 1707
T + + G+ S++ D+ +K R ++ + R +V A+GTP++N++ E+F++ +
Sbjct: 1061 TNMSTLKGVDPNQSQRADLYVKSFRFETINPGRALVLASGTPITNTLGEMFSVQRLMGH 1120
- 40
Query: 1708 DVLERYLVSNFDSWVGAFNGNIENSMELAPTGDQYQPKRFFKFNLPPELMRIYKETADI- 1766
LE + FD+W FG+ +EL P+G KY+P RF FVN+PEL+ +++ AD+
Sbjct: 1121 AALEERGLHEFDAWASTFGDTTTELELQPSG-KYKPVSRFASFVNVPPELIAMFRSFADVV 1179
- 45
Query: 1767 ---QTSMDMLDLP-VPEAKIIAVESELTAQKYYLEELVKRSDAIKSGS--VDPSRDNMLK 1820
+ + +P + + V S+ TQA K++ L +R AI+ P D +L
Sbjct: 1180 MPADLREYVKVPAISTGRRQIVTSKPTQAFKHHQMVLAERIKAIEERERPPQPQDDILLS 1239
- 50
Query: 1821 ITGEARKLAIDMRLIDPTYLSDNQKILQVVDNVERIYRDGAGDK-----AT 1867
+ + R AID+RL+D + K+ +V N RI++ AG A
Sbjct: 1240 VITDGRHAAIDLRLVDADNDNEPDNKLNNLVSNAPRIWKATAGSVYL RHDSKPFEPVPGAA 1299
- 55
Query: 1868 QMIFSDIGTPK-SKEEGFDVYNELKDLFVDRGIPKEEIAFVHDANTDEKKNLSRKVNSG 1926
QMIFSD+GT K GF Y ++D + G+P EIAF+ D E K L V +G
Sbjct: 1300 QMIFSDLGTISVEKTRGFSAYRWIRDELIRLGVPAEIAFMQDFKKSEAKQRLFGDVRAG 1359
- 60
Query: 1927 EVRILMASTEKGGTGLNVQSRMKAVHYLDVPWRPSDIVQRNGRLIRQGNMHQEVDIYHYI 1986
VR L+ S+E GTG+NVQ R+KA+H+LDVPW PS I QR GR++RQGN H EVDI+ Y
Sbjct: 1360 RVRFLIGSSETMGTVGVNVQLRLKALHHLDPVWLP SQIEQREGRIVRQGNQHDEVDIFAYA 1419
- 65
Query: 1987 TKGSFDNYLWQTQENKLYITQIMTSKDPVRSADIDE-QTMTASDFKALATGNPYLKLK 2045
T+GS D +WQ E K ++I ++ +R EDI E Q + KA+A+G+ L K
Sbjct: 1420 TEGSLDATMWONNERKARFIAAALSGDTSIRRLLEDIGEGQANQFAMAKAIASGDQRLMQK 1479
- Query: 2046 MELENELTVLENQKRAFNRSKDEYRHTISYSEKHLPIMEKRLSQYDKDIAQSLATKSQDF 2105
LE ++ LE + A + R + +E+ + + +R+++ +DI + + T +DF
Sbjct: 1480 AGLEADIAERLERLRAAHIDDQHAVRRQLRDAERDIEVSTRRIAETIGQDITRLVPTTGEDF 1539
- Query: 2106 VMRFNDQAMDNRAEAGDYLRK-LITYNRSETKEVRTLASFRGFDLKM-TTRGASEPLPET 2163
M + R EAG L K ++T + + +AS GF+L+ R + T
Sbjct: 1540 TMTVAGKDYSERKEAGRALMKEILTIVQLSPEGEAVIASIGGFELEYHGQRYGKDGRYT 1599
- Query: 2164 ISLMIVGDNQYTVALLDLK-SDVGTIQRISNAIDHIIDDQEKTOELVKDLKDLRVAKVEV 2222
L G + Y + L + + +G + R+ +A+D ++E+ ++ + D + +L +
Sbjct: 1600 TMLKRTGAD-YEIELPVTVTPLGAVSRLEHALDDFDGERERYRQRLGDARRRLASYQSRG 1658

-2129-

Query: 2223 DKVFPKEEDYQLVKAKYDVLAPLVEKEAEIEEIDAALA 2260
 + + + + L EK ++ E++ ALA
 Sbjct: 1659 E-----GSEFAFAGELAEKHRQLAEVETALA 1684
 Identities = 99/271 (36%), Positives = 153/271 (55%), Gaps = 10/271 (3%)

Query: 607 RDKVETNIVAIRLVKNLEVEHRNASPSEQELLAKYVGWGG--LANEFD-----DYNPKF 659
 +D+ NI AIRL +E R A+ EQE L ++ G+G LAN F ++ +
 Sbjct: 80 KDRARDNIAAIRLAAEIEASERPATREEQETLIRFTGFGASDLANGVFRRPGELEFRKGW 139

Query: 660 SKEREELKSLVTDKEYSDMKQSSLTAYYTDPSLIQMWKLERDGFSGKILDPSMGTGN 719
 + +L+ V + +Y+ + + + A++T ++R +W L+R G+ GG++L+P +GTG
 Sbjct: 140 DEIGSDLEDAVGETDYASLARCTQYAHFTPEFIVRAIWSGLQRLGWRGGRVLEPGIGTGL 199

Query: 720 FFAAMPKHLREKSELYGVELDTITGAIKHLHPNSHIEIKGFETVAFNDNSFDLVISNVP 779
 F A MP+ LR+ S + GVELD +T I + L P + I F SFDL I N P
 Sbjct: 200 FPALMPEALRDLSHVTGVELDPVTACIVRLQPRARILTGFARTEL-PASFDLAIGNPP 258

Query: 780 FANIRIADNRYDRP--YMIHDYFVKKSLDLLHDGGQVAIISSGTMDKRTENILQDIRET 837
 F++ + +R R +HDYFV +S+DLL G A ++S+GTMDK Q I T
 Sbjct: 259 FSDRTVRSRAYRSLGLRLHDYFVARSIDLLKPGAFAAFVTSSGTMDKADSAARQHIATT 318

Query: 838 TEFLGGVRLPDSAFKAIAGTSVTTDMLFFQK 868
 + + +RLP+ +F+A AGT V D+LFF+K
 Sbjct: 319 ADLIAAIRLPEGSFRADAGTDVVVDILFFRK 349

SEQ ID 5856 (GBS327N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 8-10; MW 140kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 11-13; MW 115kDa) and in Figure 182 (lane 8; MW 115kDa).

Purified GBS327N-GST is shown in Figure 243, lane 5; Purified GBS327N-His is shown in Figure 235, lane 5.

GBS327C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 14; MW 73kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1882

A DNA sequence (GBSx1990) was identified in *S.agalactiae* <SEQ ID 5857> which encodes the amino acid sequence <SEQ ID 5858>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

| |
|---|
| bacterial cytoplasm --- Certainty=0.3656(Affirmative) < succ> |
| bacterial membrane --- Certainty=0.0000(Not Clear) < succ> |
| bacterial outside --- Certainty=0.0000(Not Clear) < succ> |

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2130-

Example 1883

A repeated DNA sequence (GBSx1991) was identified in *S.agalactiae* <SEQ ID 5859> which encodes the amino acid sequence <SEQ ID 5860>. This protein is predicted to be giant membrane protein. Analysis of this protein sequence reveals the following:

```

5      Possible site: 33
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:AAG19662 GB:AE005054 calcium-binding protein homology; Cbp
      [Halobacterium sp. NRC-1]
      Identities = 22/43 (51%), Positives = 29/43 (67%), Gaps = 1/43 (2%)

      Query: 9   KDSQDQGLTDAQELAL-GTDPQSVDTGDDGQADLEELQSGHSP 50
                +D+D DGL+D E+ + GTDP DTDGDG D EL++G P
20      Sbjct: 198 RDTDDGLSDGVEVRVAGTDPFTRDTGDDGVDDAAELRAGSLP 240

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1884

A DNA sequence (GBSx1992) was identified in *S.agalactiae* <SEQ ID 5861> which encodes the amino acid sequence <SEQ ID 5862>. Analysis of this protein sequence reveals the following:

```

30      Possible site: 52
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.39    Transmembrane 1609 -1625 (1609 -1625)
      INTEGRAL    Likelihood = -1.81    Transmembrane 30 - 46 ( 29 - 46)

      ----- Final Results -----
35          bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

LPXTG motif 1600-1604

The protein has homology with the following sequences in the GENPEPT database.

```

40      !GB:X57841 antigen I /II [Streptococcus sobrinus] (v...

      >GP:CAA40973 GB:X57841 antigen I /II [Streptococcus sobrinus]
      Identities = 419/1436 (29%), Positives = 608/1436 (42%), Gaps = 310/1436 (21%)
45      Query: 23   KSKKYRTLCSVALGTMVTAVVWGGTVAHADEVTTTSV----DTTIQRTQ--NPATNLPEA 76
                  K K RTL LGT + A A G A A+E +T+ DT + TE NPATNLP+
      Sbjct: 23   KVKSGRTLSGALLGTAILASGA--GQKALAEETSTTSTSGGDTAVVGTETGPNPATNLDPK 80

      Query: 77   QPNP-----VSEQTESMASTGQSGAIAVTVPHDTVT-----QAVE 112
                  Q NP V T + +S VTV D + +
      Sbjct: 81   QDNPSQAETSQAQARQKTGAMSVDVSTSELDEAAKSPQEGVTVSQDATVNGKTVPEPSD 140

      Query: 113  EAKAEGVSTVEDSPMDLGNTRSAVET-----NQQIS-----K 144
                  EA + +D + + A E NQ+I+ K
55      Sbjct: 141 EANQKEPEIKDDYSKQADIQKATEDYKASVAANQAETDRINQEI AAKKAQYEQDLAANK 200

```

-2131-

5

Query: 145 AD-----ADTQKQVETINEVTK----TYKADKATYESNKARIEQEN 181
 A+ A QK + I + Y A K Y+ AR++ N
 Sbjct: 201 AEVERSLMRMRKPRPIYEAKLAQNQKDLAAIQQANSQSQAAYAAKEAYDKEWARVQAAN 260

10

Query: 182 KELSQAAYEGANQTKGETNAWVDTKVNDLKARYADADVTVKEQ-----VVSSGNGTSVL 234
 +AYE A N + ++ ++ R A AD K +GN +
 Sbjct: 261 AAKKAYEEALAANTAKNDQIKAEIEAIQORSADYEAQYKDLAAAQAGNAANE 320

15

Query: 235 DY----TNYGKAVETIQSTNEQAVADY----LTKKTKADDIVAKNQAIQKENE----- 280
 DY Y + + +Q+ N A Y K I A+N+AIQ+ +A
 Sbjct: 321 DYQAKKAAYEQELARVQAANAAKQAYEQALAANSKNAQITAENEAIQQNAQAKADYEA 380

20

Query: 281 -----GLANAKADNEAIERRNQAGQAAVDAEN---RAGQAAVDQANQEKQQLVSDRAA 330
 LA A++ N A E Q AA + E +A AA QA +++ Q + + A
 Sbjct: 381 KLAQYQKDLAAQSGNAANEADYQEKLAAYEKELARVQAANAAKQAYEQVQVQANAKNA 440

25

Query: 331 EIEAITKRNKEKEAAARKENEIDAINTKEMERYQORDLAIS----- 372
 EI + +E+ A A+ + E + +E+ +Y++DLAE
 Sbjct: 441 EITEANRAIRERNAKAKTDYELKSKYQEEELAQYKKDLAEYPAKLQAYQDEQAAIKAALA 500

30

Query: 373 -----KGEEGYISEALAQALNNGEPQAOHGAI TRN----- 404
 K E+G +SE AQ+L + + EP AQ +T
 Sbjct: 501 ELEKHKNEDGNLSEPSAQSLS-VYDLEPNAQVALVTDGKLLKASALDEAFSHDEKNYNNHL 559

35

Query: 405 --PDQI-----ISTGDALLGGYSRILDSTGF-----FVYDMFKTGETLS 441
 PD + +++ L G + D G+ F + K G++ +
 Sbjct: 560 LQPDNLNVTYLEQADDVASSVELFGNFG---DKAGWTTTVSNGAEVKFASVLLKRGQSAT 616

40

Query: 442 FNYQNLQHARFDGKKISRVTYDITNLVSPAG-----TNAVKLVPNDPTEGFIAYRNDGN 496
 Y NL+++ ++GKKIS+V Y T V P T V L + DPT G A G
 Sbjct: 617 ATYTNLKNSYNGKKISKVVYKYT--VDPDSKFQNPFTGNVWLGIPTDPTLVGFASAYTGQ 674

45

Query: 497 GDWRD---KMEFRVVAKYYLEDGSQVTFSEKPGVFTHSSLNHNNDIGLEYVKDSSGKFV 553
 + T K EF +Y EDG+ + F + + +SLN +E KD SG FV
 Sbjct: 675 NEKDTSIFIKNEF---TFYDEGDNPIDFDN---ALLSVASLNREHNSIEMAKDYSGTFV 727

50

Query: 554 PINGSTVQVTN-----EGLARSLGSNRASDLNLPEEWDTTSSRYAYKGAIV 599
 I+GS++ N EG + RAS+ WD+ + ++ GA
 Sbjct: 728 KISGSSIGEKNGMiyATDTLNFKKEGGSLHTMYTRASEPG--SGWDSADAPNSWYGAGA 785

55

Query: 600 STVTSNGTY-----TVTFGQGDMPQNVGL-----SYWFALN----- 630
 ++ N Y T +MPQ G + W++LN
 Sbjct: 786 VRMSGPNNYITLGATSATNVLSLAEMPQVPGKDNTAGKKPNIWYSLNGKIRAVNVPKVTK 845

50

Query: 631 --TLPVARTVTPYSPKPHVTVEL-----EPIPEPITVTPDIYTPKFTTPEKPVFTFT---- 679
 P P P V EL EP EP TP P PEKPV T
 Sbjct: 846 EKPTPPVEPTKPEPTYEVEKELVDLPVEPKYEP-EPTPPSKNPDQSIPEKPVETTYEVE 904

60

Query: 680 ----PKPLDEVVQPSLTTLTKVT-----LPVKPIPKELTPP-----QVPTV 716
 P P++ + T + T PV+P + LPTPP VPTV
 Sbjct: 905 KELEPAPEPSYEKEPTPPQSTPDQEEPEKPVPSYQSLPTPPVEPVYETVPGPVSVPTV 964

65

Query: 717 HYHAYRLTTTSEIMKEVVNSDQANLHEKTVAKDSTVIYPLTVDALSPNRAQTTSLIFEDY 776
 YH Y+L + KE+ N D ++ + VAK STV + L L R +TTS + D
 Sbjct: 965 RYHYKLAQVPGVTKEIKNQDDLIDKTLVAKQSTVKFQLKTADLPAGRPETTSFVIMDP 1024

Query: 777 LPAGYLFDKETQKENGNYVLSFDETKNFVTLTAKENLLQEVNKDLTQVYQLTAPKLYGS 836
 LP+GY + E T+ + + S+D + VT TA L +N+DLT+ P + G
 Sbjct: 1025 LPSGYQLNLEATKVASPGFEASYDAMTHTVTFATAETLAALNQDLTKAVATITYPTVVQG 1084

Query: 837 VQNDGATYSNSYKLLLNKGTNNAYTVTSNVVTVRTPG-----DGETTTLITPDKNENAD 891
 V NDGATY+N++ L++N +AY + SN+V V TPG D + ITP K N+N +
 Sbjct: 1085 VLNDGATYTNFTLMVN----DAYGIKSNIVRVITPGKPNPDNPSNNYITPHKVNKNEN 1140

Query: 892 GVLINDTVVALGTTNHYRLTWDLQYKGDERSAKETIARGFFVDDYPPEVLDVVENGTAI 951
 GV+I+ V GTTN+Y LTWDLQYKGD+SAKE I +GFF+VDDYPPEE LD+ + +
 Sbjct: 1141 GVVIDGKSVLAGTTNYYELTWDLQYKGDKSAKEIIQKGFVDDYPPEALDLRTDLIKL 1200

-2132-

Query: 952 TTLDGQKVSIGITVKNYASLNEAPKDLQDKLARAKITPTGAFQVFMDDNQAFYDQYVQTG 1011
 T +G+ V+G++V +YASL AP +QD L +A I P GAFQVF DD QAFYD YV TG
 Sbjct: 1201 TDANGKAVTGVSVADYASLEAAPAAVQDMLKKANII PKGAFQVFTADDPQAFYDAYVVTG 1260

Query: 1012 TSLALLTKMTVKDSLYGQTKTYINKAYQVDFGNGYETKEVTNTLVSPPEKKQ-NLNKDKV 1070
 T L ++T MTVK + +Y N+AYQ+DFGNGYE+ V N + P+K L D
 Sbjct: 1261 TDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVPKINPEKDVTLTMDPA 1320

Query: 1071 D---INGKPLVGTQNHYTLSWDLQYRGIKADNSQIAQGFYFVDDYPE-----EALLPD 1122
 D ++G+ + + +Y L + I AD+++ + F DDY +
 Sbjct: 1321 DSTNVDGQTIALNQVFNYRLIGGI-----IPADHAEELFEYSFSDDDYDQTGDQYTGQYKA 1375

Query: 1123 EAAIQFVTS DGKTV-SGITVKS--SQLLEAPKTLQAAFSKQKIOPKGAFQVFMPE 1175
 A + DG + +G + SY +Q+ EA + F + ++ F E
 Sbjct: 1376 FAKVDLTLKDGTLTIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAE 1431
 Identities = 209/442 (47%), Positives = 280/442 (63%), Gaps = 27/442 (6%)

Query: 1198 TVLETMLNSGKSY-ENVAYQVDFGQAYETNTVINFVPK-----VTPHKSNTNQ 1244
 TV+ +LN G +Y N V+ ++N V P +TPHK N N+
 Sbjct: 1080 TVVGQVLNDGATYTNNTFLMVNDAYGIKSNIVRVTTPGKPNPDNPSNNYITPHKVNKNE 1139

Query: 1245 EGISIDGKTIVLPNTVNYKIVLDYSQYKDMVVTDVLAKGYMDDYPEEALTLNPDGIQ 1304
 G+ IDGK+VL T NYY++ D QYK +++ KGF+ VDDYPEEAL L D I+
 Sbjct: 1140 NGVVIDGKSVLAGTTNYYELTWDLQYKGDKSAKEIIQKGGFFVDDYPEEALDLRTDLIK 1199

Query: 1305 VLDKDGNRVSGISVSTYASLSEAPKVVQDAMAKRQFTPKGAIQVLSDDPKVFYDVTYVKT 1364
 + D +G V+G+SV+ YASL AP VQD + K PKGA QV ++DDP+ FYD YV T
 Sbjct: 1200 LTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANII PKGAFQVFTADDPQAFYDAYVVT 1259

Query: 1365 GQTLVVTLPMPTVKNELTKTGGQYENTAYQIDFGLAYVTETVNNVVPKLDPOKDVVIDLSH 1424
 G L + PMTVK E+ KTGG YEN AYQIDFG Y + VNNVVPK++P+KDV + +
 Sbjct: 1260 GTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVPKINPEKDVTLTMDP 1319

Query: 1425 KDA-SLDGKEVALHQTFNYRLVGAMIPSNRATDLFEYGFEDNYDEKHDEYNGVYRSYLMT 1483
 D+ ++DG+ +AL+Q FNYRL+G +IP++ A +LFEY F D+YD+ D+Y G Y+++
 Sbjct: 1320 ADSTNVDGQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDDYDQTGDQYTGQYKAFKV 1379

Query: 1484 DVILKDGSVLKEGTEVTKYTLQQVDTENGLVSI SFDKSFLETVSDDSAFQADVYLQMKRI 1543
 D+ LKDG+++K GT++T YT QVD NG + ++F + FL +VS DSAFQA+VYLQMKRI
 Sbjct: 1380 DLTLDKGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAEVYLQMKRI 1439

Query: 1544 AAGQVENTYLHTVNGYVISSNTVVTHTPQPEEPSNPQ-----TPQPPIETIEPPV 1595
 A G NTY++TVNG SSNTV T TP+P++PSP P P Q PP
 Sbjct: 1440 AVGTAFANTYVNTVNGITYSSNTVRTSTPEPKQPSVPDPKTTTIVVFQPRQKAYQPAPPA 1499

Query: 1596 PASILPNTGEQES----LLGLI 1613
 A LP TG+ + LLGL+
 Sbjct: 1500 GAQ-LPATGDSSNAYLPLGLV 1520
 Identities = 100/210 (47%), Positives = 137/210 (64%), Gaps = 4/210 (1%)

Query: 1060 PKKQNLNKKVDINGKPLVGTQNHYTLSWDLQYRGIKADNSQIAQGFYFVDDYPEEAL 1119
 P K N N++ V I+GK +L GT N+Y L+WDLQY+G K+ I +GF++VDDYPEEAL
 Sbjct: 1132 PHKVNKNENGVIDGKSVLAGTTNYYELTWDLQYKGDKSAKEIIQKGGFFVDDYPEEAL 1191

Query: 1120 LPDEAAIQFVTS DGKTVSGITVKSYSQLEAPKTLQAAFSKQKIOPKGAFQVFMPEDPQA 1179
 I+ ++GK V+G++V Y+ L AP +Q K I PKGAFQVF +DPQA
 Sbjct: 1192 DLRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANII PKGAFQVFTADDPQA 1251

Query: 1180 FFESYVTKGENITIVTPMTVLETMLNSGKSYENVAYQVDFGQAYETNTVINFVPKVTPHK 1239
 F+++YV G ++TIVTPMTV M +G SYEN AYQ+DFG YE+N V N VPK+ P K
 Sbjct: 1252 FYDAYVVTGTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVNNVVPKINPEK 1311

Query: 1240 SNT----NQEGISIDGKTIVLPNTVNYKIV 1265
 T + ++DG+T+ N V Y+++
 Sbjct: 1312 DVTLTMDPADSTNVDGQTIALNQVFNYRLI 1341

There is also homology to SEQ ID 598.

-2133-

SEQ ID 5862 (GBS76) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 2; MW 17.4kDa). The GBS76-His fusion product was purified (Figure 196, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 294), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1885

- 10 A DNA sequence (GBSx1993) was identified in *S.agalactiae* <SEQ ID 5863> which encodes the amino acid sequence <SEQ ID 5864>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEi). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15      bacterial cytoplasm --- Certainty=0.2765(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

- 20 A related GBS nucleic acid sequence <SEQ ID 9931> which encodes amino acid sequence <SEQ ID 9932> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB52382 GB:U36837 AbiEi [Lactococcus lactis]
Identities = 51/206 (24%), Positives = 90/206 (42%), Gaps = 23/206 (11%)

25 Query: 17 KNNIGIVTNKDCKALGIPTIYLTREKEGIIFRVEKGIFLTQNGDYDEYYFFQYRFPKAIF 76
      K G + K + GI YL + + + V+KG+++ + D + FQ ++ KA+
Sbjct: 76 KYKGNIRKIVRDEGISDYLRKFLKYNLTEVDKGVYIFPHKKKDSLFIQFQKYKSAVI 135

Query: 77 SYISALYLQOFTDEIPQYFDVTVPRGYRF-----NTPPANLNI 114
30 S+ ++LYLQ D IPQ ++VP Y N N+ I
Sbjct: 136 SHETSLYLQDVIDYIPQKIQMSVPEKYNISRIQEPHENRLTSYNYVDINSNNIMDKNIPI 195

Query: 115 HFV-SKEYSELGMTTVPTPMGNNVRVYDFERIICDFVIHREKIDSELFVKTLQSYGNYPK 173
      + V +K S + TV + +G +RV R I D + K + E+ + ++ Y
35 Sbjct: 196 NLVRNKSISPTQIETVNSFLGLPLRVSTIARSIVDVLKPSHKAEEVKEQAIKYYLERFP 255

Query: 174 KNLAKLYEYATKMTLEKVKQTLEVL 199
      N+ +L A N L++++ L +L
40 Sbjct: 256 DNIVRLKRIAKTQNVLKELEYLILL 281
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1886

- 45 A DNA sequence (GBSx1994) was identified in *S.agalactiae* <SEQ ID 5865> which encodes the amino acid sequence <SEQ ID 5866>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEii). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
50      INTEGRAL Likelihood = -1.12 Transmembrane 260 - 276 ( 259 - 277)
```

-2134-

----- Final Results -----

bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB52383 GB:U36837 AbiEii [Lactococcus lactis]

Identities = 76/276 (27%), Positives = 135/276 (48%), Gaps = 19/276 (6%)

Query: 14 SKNTGLTFNSVMTYYFLEVILKKLSQSSYSNHYIFKGGFLLSNVIGVESRSTVDIDFLFH 73
 ++N + + Y E L +LS S Y ++ KGGFL+ + R+T D+D
 Sbjct: 12 TRNDDIGIENYRIRYATERFLTRLASQYKEKPVFKGGFLIGVTYNLSQRTTKDLDTALI 71

Query: 74 QITLSEETVKQQLKEIL-ADSEEGISFVIQSITTIKESDDYGGYRATISCQLE--NIKQV 130
 +++++ + EI D E+ + F ++ +T+ ++ Y GYRA + N +
 Sbjct: 72 DFKSDAQSIERVITEICNIDLEDQVLFLKELTSSQDMRIYPGYRAKLMMFPDGNTRID 131

Query: 131 IHLDIATGDVVTQPITYDYKAIFDE----DNFPILAYTIETILAEKLQTIYSRNFLNS 185
 LDI GD +TP+ IF+E ++AY ETI AEKL+TI +R +N+
 Sbjct: 132 FDLDIGVGDRITPEAKKIKIPLIFNEVKGVEKQIEVLAYPKETIQAEKLETILTRGVNT 191

Query: 186 RSKDFYDVYIL--SKLKKKDIDFNQLKNACQRTFSYRE-TELDFEKIIE-----LLERFK 237
 R KD+YD ++L + I F A + T+ +R T+ E++ E L E +
 Sbjct: 192 RMKDYDFHLLLTQENSNSISIFY---AFKNTWEFRNPTQFIDEELFEDWLFILDEILE 248

Query: 238 SDPTQNQQWQNYSKKYSYTKGISLANVLDEMISLIT 273
 S + + W NY K +Y K +++ ++ E+ ++
 Sbjct: 249 SKELKEKYWPNIYKDRNYAKHLNMDDIISEIKEFVS 284

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1887

A DNA sequence (GBSx1995) was identified in *S.agalactiae* <SEQ ID 5867> which encodes the amino acid sequence <SEQ ID 5868>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1137(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1888

A DNA sequence (GBSx1996) was identified in *S.agalactiae* <SEQ ID 5869> which encodes the amino acid sequence <SEQ ID 5870>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

-2135-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2782(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1889

A DNA sequence (GBSx1997) was identified in *S.agalactiae* <SEQ ID 5871> which encodes the amino acid sequence <SEQ ID 5872>. Analysis of this protein sequence reveals the following:

15 Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.14 Transmembrane 310 - 326 (301 - 334)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAG38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
 Identities = 272/344 (79%), Positives = 307/344 (89%)
 Query: 568 VYVNPAFYFPKVIQVQTILPTIGQFGGDEFERAKAIYDYLKSKGATNQAI AAILGNWSV 627
 +YVNP FYFPKVIQ+QTILP IGQFGGDEFERAK IY++LKS+GA+ QAIAAILGNWSV
 30 Sbjct: 1 MYVNPQFYFPKVIQLQTILPAIGQFGGDEFERAKHIYEFLKSQGASFPQAIAAILGNWSV 60
 Query: 628 ESSINPKRAEGDYLSPFVGATDSSWDEGWLTNGPTIYNGRYPNILKRLGLGQWTDTA 687
 ESSINPKRAEGDYL+PPVG WDDE WL + GP IY+G YPNIL RGLGLGQWTDTA
 35 Sbjct: 61 ESSINPKRAEGDYLTPFVGVPPIPPWDESWLAIGGPATYSGAYPNILHRLGLGQWTDTA 120
 Query: 688 DGSRRHTLLLEYAKGKHQKWDYDLGLQLDFMLYGDSPYYTNWLKDFFKNSGSPASLAQLFL 747
 DGS RHT LL YA+ +++KWYDL LQLDFML+GDSPYY +WLKDFFKN+GS A+LAQLFL
 40 Sbjct: 121 DGSTRHTALLNYARTQNKWYDLDLQLDFMLHGDSPYYQSWLKDFFKNTGSAANLAQLFL 180
 Query: 748 IYWEGNSGDKLLERQTRASEWYYQIEKGFSQPNGGTAQSDPKALEAVREDLFENSIPGGG 807
 YWEGNSGDKLLERQTRA+EWYYQIEKGFSQ NGG A+SDP++LE VR DL+++S+PGGG
 45 Sbjct: 181 TYWEGNSGDKLLERQTRATEWYYQIEKGFSQTNGGQAKSDPQSLEGVRGDLYDHSVPGGG 240
 Query: 808 DGMGYAYGQCTWGVAARINQLGLKLKNGEKIPIIISTMGNGQDWVRTAASLGGETGTSP 867
 DGM YAYGQCTWGVAAR+NQLGLKLKNGEKI II+TMNGQDWV T++SLGGETG++P
 50 Sbjct: 241 DGMAYAYGQCTWGVAARMNQLGLKLKGRNGEKISIINTMGNGQDWVATSSSLGGETGTSP 300
 Query: 868 QEGAILSFAGGGHGTPTTEYGHVAFVEKVYPDGSFLISETNYNGN 911
 + GAI+SF GG HGTP YGHVAFVEKVY DGSFL+SETNY GN
 Sbjct: 301 RAGAIVSFVGGTHGTPASYGHVAFVEKVYDDGSFLVSETNYNGN 344

SEQ ID 5872 (GBS74d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 3 & 4; MW 95.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 5-7; MW 70.5kDa) and in Figure 179 (lane 9; MW 70.5kDa).

55 GBS74d-His was purified as shown in Figure 233, lane 7-8.

-2136-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1890

A DNA sequence (GBSx1998) was identified in *S.agalactiae* <SEQ ID 5873> which encodes the amino acid sequence <SEQ ID 5874>. This protein is predicted to be TrsE-like protein. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38042 GB:AF295925 Orf26 [Streptococcus pneumoniae]
Identities = 618/782 (79%), Positives = 712/782 (91%), Gaps = 1/782 (0%)

Query: 1 MKKLKHSMSK-TSSNDKKQKTKTKQKEISPSTVNTLAYQGLFQNGLMQVSPSYFSQTYL 59
MK+ +++K + TS+ +KK++ K +K+E+ PST NTL+YQ L+QNGLMQV YFSQ+YL
Sbjct: 3 MKRKSNTLKKQQTSTTNKKEEVKDKKEEVLPTANTLSYQALYQNGLMQVKEDYFSQSYL 62

Query: 60 LGDVNYQTVGLDDKGAIVEKYSDLI NSLDDKTNFQLTIFNQVNLEKFRKSILYPLQEDG 119
LGDVNYQTVGL+DKGAI+EKYSDLI SLDD+TNFQLTIFN+++NLEKFR S+LY +EDG
Sbjct: 63 LGDVNYQTVGLEDKGAIIEKYSDLIKSLDDQTNFQLTIFNKRINLEKFRHSVLYEEKEDG 122

Query: 120 FDIYRDELNRMMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEIFKSGFSEID 179
+D+YR ELNRMM+ NL++GENNFSAVK +SFG+ D PK A+RSLSQIGEIFKSGFSEID
Sbjct: 123 YDSYRKELNRMMNQNLDSGENNFSAVKLISFGRKDSNPKQAYRSLSQIGEIFKSGFSEID 182

Query: 180 VSLGLLGGEERVNVLADMLRGENHLPFSYKDLTSLSGQSTKHFIAPTYLSFKHKNHIELD 239
L GEERVN+LADMLRGE+HLPPSY+DLT SGQ+T+HFIAP L FK+KN++++D
Sbjct: 183 ARFESLAGEERVNLLADMLRGEHHLPPSYRDLTRSGQTTTRHFIAPNLLDFKNKNYLQIND 242

Query: 240 RLLQIVYVRDYGMEIGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKI 299
RLLQIVYVRDYGMEIGD+FIRDLMQ DLE+++SLHA+ STKS+ M KLRTKKTLMESQKI
Sbjct: 243 RLLQIVYVRDYGMEIGDQFIRDLMQGDLELIVSLHAQSSTKSDAMKKLRTKKTLMESQKI 302

Query: 300 GEQQKMARTGIYLEKVGHVLENNIDEAEALLQMTQTGDKLPDFTVFLIGVLADTEDQLKQ 359
GEQQK+ARTGIYLEKVGHVLE+NIDEAE LL+TMT+TGDKLF TVFLIGV E++LKQ
Sbjct: 303 GEQQKLARTGIYLEKVGHVLESNIDEAEELLKMTTETGDKLPQTVFLIGVFGQDEEELKQ 362

Query: 360 SLDIIKQVAGSNDMIIDNLTMYQEAFAFNSLLPFGKNYLEGVSRSLTNSIAVNAPWTSVD 419
+LD ++QVAGSND++ID L YMQEAFAFNSLLPFG ++LEGVSRSLTNSIAVN+PWTSVD
Sbjct: 363 ALDTVQQVAGSNDLMIDKLPYMQEAFAFNSLLPFGCDFLEGVSRSLTNSIAVNSPWTSVD 422

Query: 420 IHDKGGKFYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSD 479
+ D+ GK+YGINQISSNII+IDR LNTPSGLILGTSGAGKGMATKHEIITK+KE+ +
Sbjct: 423 LQDRSGKYYGINQISSNIITIDRSLNTPSGLILGTSGAGKGMATKHEIITTKIKESGEN 482

Query: 480 TEIIIVDPENEYSIIGQAFGESIDIAPDSTTFINVLLESDENMDEDPVKVSEFLLSWI 539
TEIIIVDPE EYS+IG+ FGGE IDIAPDS T+LNVL+LS+ENMDEDPVKVSEFLLS+I
Sbjct: 483 TEIIIVDPEAEYSVIGRTFGGEMIDIAPDSEYTLNVLDLSEENMDEDPVKVSEFLLSFI 542

Query: 540 GKLLDRKMDGREKSLIDRVTRLTYKHFDTPSLVEWVFLVLSQQPEQEAQDLALDMELYVEG 599
GKLLDRKMDGREKS+IDRVTRLTY+ F PSL EWVFLVLSQQPE+EA++LALDMELYVEG
Sbjct: 543 GKLLDRKMDGREKSIIDRVTRLTYQSFKEPSLEEWVFLVLSQQPEEEAQNALALDMELYVEG 602

Query: 600 SLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLGKKTWIFYD 659
SLDIFSH+TNI+T S+FLIYNVKKLGDELKQIALMV+FDQIWNRVV+NQKLGKKTWIFYD
Sbjct: 603 SLDIFSHKTNIQTGSNFLIYNVKKLGDELKQIALMVVFDQIWNRVVRNQNQKLGKKTWIFYD 662

-2137-

Query: 660 EMQLLLLDKYASDFFFKLWSRVRYGAIPGTGTONVETLLLDANGRRRIANSEFMILLKQ 719
 E++LLLLDKY SDDFFKLWSRVRYGA PTGTONVETLLLD NGRRIANSEFMILLKQ
 Sbjct: 663 EIELLLLDKYPSDFFFKLWSRVRYGASPTGTONVETLLLDPNRRRIANSEFMILLKQ 722

5 Query: 720 AKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLFDMISTDPE 779
 AK+DREELV +LGLSKELEKYLVNPEKGAGLIKAGS VVPFKNKIPQ ++LFDIM +DP+
 Sbjct: 723 AKNDREELVQLLGLSKELEKYLVNPEKGAGLIKAGSVVPFKNKIPQGSQFLDIMRSDPD 782

10 Query: 780 KM 781
 KM
 Sbjct: 783 KM 784

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8925> and protein <SEQ ID 8926> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -26.26
 GvH: Signal Score (-7.5): -3.87
 Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 0 value: 6.26 threshold: 0.0
 PERIPHERAL Likelihood = 6.26 335
 modified ALOM score: -1.75

25 *** Reasoning Step: 3

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear)

The protein has homology with the following sequences in the databases:

35 33.5/57.2% over 789aa

Enterococcus faecalis

GP|8100663| TrsE-like protein Insert characterized

ORF01332(319 - 2628 of 2949)

40 GP|8100663|gb|AAF72347.1|AF192329_8|AF192329(2 - 791 of 799) TrsE-like protein
 {Enterococcus faecalis
 }

%Match = 20.7
 %Identity = 33.4 %Similarity = 57.2
 45 Matches = 259 Mismatches = 323 Conservative Sub.s = 184

| | | | | | | |
|---|-----|-----|-----|-------------------------------------|---------------|-----|
| 210 | 240 | 270 | 300 | 330 | 360 | 387 |
| SCYLGSIAPTIIYHLKYTSSTVFIMN*RCQTAHLLEEKETNVKKLKHSMKSKTSSNDKKQKTKTKQKEI-----S | | | | | | |
| | | | | | : : : : | |
| | | | | MSKKEIPRETEKTKLTRAQRKEIDAVIRKYKGDGR | | |
| | | | | 10 | 20 | 30 |

| | | | | | | | |
|--|---------------|-----|-----------|-----|-----|-----------|-----|
| 414 | 444 | 474 | 504 | 534 | 564 | 594 | 624 |
| PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDVNYQTIVGLDDKGAIVEKYSIDLINSLDDKTNFQLTIFNQKVNLEKFR | | | | | | | |
| | : : : : | | : : : | | | : : : | : |
| PHTAQQSIPYEVVMDGVCVRVSPGVFSKCI EFADISYQLAQPDQTATIFEKLCILYNYVDASIHQFSFLNRKVDPVQYA | | | | | | | |
| 50 | 60 | 70 | 80 | 90 | 100 | 110 | |

| | | | | | | | |
|---|-----|-----|-----|-------|-----|-----|-----|
| 654 | 684 | 714 | 744 | 774 | 804 | 834 | 864 |
| KSILYPLQEDGFDTYRDELNRMMDANLEAGENNFSVAVKFLSFGKSDQTPKLAFRSLSQIGEIFKSGFSEIDVSLGLIGGE | | | | | | | |
| | | : | | : : | : | | : |
| KSFEIAPQGDGDDFDDIRAETYGILQKQLANGNNGMVKTKYLTFTIEAESVKAARLRKRIGFDLLGYFKSMGAVAHVMDGW | | | | | | | |

130 140 150 160 170 180 190

894 921 951 981 1011 1041 1071 1101

ERVNVVLADMLRGENHL-PFSYKDLTTLSGQSTKHFIAPTLYLSFKHKNHIELDDRLQLIVYVRDYGMELGDKFIRDLMQSDL
|:|:| : : : |:-| ||||| |:| : : : : || | : |: :
ERLNLHLGVYHPDGEIFNFDWKWLAPSGSLSTKDFIAPSSLCPGNKTFGMGGKYGAVSFLQILSPELSDMLADFLNTES
210 220 230 240 250 260 270

1131 1161 1191 1221 1251 1281 1311 1341

EVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQQKMARFTGIYLEKVGHVLENNIDEAEALLQTMQTQDGLFDTVFLI
|::|| : :: : : | | :: |||:| |:| :: : | ::|:| | : ::|| |:
GVLVNLHVQAIEQTKAIKTIKRKITDLDAKMAIEQKKA VRSGYDMDILPSDLATYGEDAKKLLTKLQTRNERLFLQTLFLV
290 300 310 320 330 340 350

1371 1401 1431 1461 1491 1521 1551

GVLADTEDQLKQSLDI IKQVAGSNDMIIDNLTMYQEAFAFNSLLPFGKNYLEGVSRSLTNSIAVNPWTSVDIHDKGGK-
:| |:| : : || : : : | | | : | |:| | : : || |:|:| | : : : |
LNVADTKQKLNNDVFQAAGVAQKHNCPLVRLDYQEQGLASSLPLGVNQI-KIQRSLTSSVAVFPVFTQELFQGGAA
370 380 390 400 410 420 430

1608 1638 1668 1698 1728 1758 1788 1818

FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDTEIIIVDPENEYSIIGQAFGGESIDIA
:| | | | | |:| |:| : |:| | | |:| |:| : | |:| | : |:| | | | : : : |:| : :
YYGINAKSRNMIMLDRKQARCPNALKLGTPGSGKSMCKSEIVSVFLITPD---DIFISDPEAEYYPLVKRLHGQVIRLS
450 460 470 480 490 500 510

1848 1875 1905 1935 1959 1989 2019

PDSTTFLNVLELS-DENMDEDPVKVKSEFLLSWIGKLLDRK--MDGREKSLIDRVTRLTYKHFDTPLSVEWVFLS----
| | |:| |:| : : |:|:| :| |:|:| : : | : : |:|:| | : |:| : |:|
PTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCELVMGCKNGLEAIEKTVIDRAVRVIYRPYLADPRPENMPILSDLHK
530 540 550 560 570 580 590

2058 2088 2118 2148 2178 2208 2238 2268

---QQPEQEAQDLALDMELYVEGSLDIFSHRTNIKTDSHFLTYNVKKLGDELQKQIALMVI FQDIQWNRVVKNGQLGKKTWI
| || :| ::|| | |:|:| |:|:| : : : : |:| |:|:| : : : | | | | :| | |
ALLDQHVPEDARVAQALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKGLMLIVQDQIWGRVTANRSQGKATWY
610 620 630 640 650 660 670

2298 2328 2358 2388 2418 2448 2478 2508

YFDEMQLLLLLDKYASDFFFKLWSRVRKYGAIPGTGITQVETILLDANGRIIANSEFMILLKQAKSDREELVHMLGLSKE
: | | |:| : : : : |:| | |:| | | |:|:| | | | |:| | | | |
FADEFHLLLKEEQTAAYS AEIWRKFRKWGGIPTGATQNVKDLLSSPEIENILENSDFTILLNQASGDRKILAE RLNLSTE
690 700 710 720 730 740 750

2538 2568 2598 2628 2658 2688 2718 2748

LEKYLVNPEKGAGL IKAGSTVVPFNKIPQHTKLFDIMSTDPEKMRT*DERG*KASQTG*AKLSKQLKISSYALSERS*D
: |:| | | |:| : |:| | |:|:| |:|:| : : :
QQKYIDNSEPGEGLLIFENVVLPFTNP IPHNTQLYKIMTTRLNEVAGV
770 780 790

This protein might be involved in vancomycin resistance

55 >GP|8100663|gb|AAF72347.1|AF192329_8|AF192329 TrsE-like protein
{Enterococcus faecalis}

Score = 427 bits (1086), Expect = e-118

Identities = 257/785 (32%), Positives = 431/785 (54%), Gaps = 28/785 (3%)

60
Query: 9 DKKQKTKTKQKEIS-----PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDV 56
+K + T+ Q+EI P T ++ Y+ ++ +G+ +VSP FS+ D+

65 Query: 57 NYQTVGLDDKGAIVEKYSDDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDGFDTY 116

-2139-

+YQ D + AI EK DL N +D + Q + N+KV+ ++ KS Q D FD
 Sbjct: 71 SYQLAQPDQTATIFEKLCDLNYVDASIHIQFSFLNRKVDPVQYAKSFETAPQGDDFDDI 130
 Query: 117 RDELNRMM DANLEAGENNFS AVKFLSPGKSDQTPKLAFRSLSQIG EYFKSGFSEIDVSLG 176
 5 R E ++ L G N K+L+F ++ K A L +IG F +
 Sbjct: 131 RAEYTGILQKQLANGNNGMVKT KYLTFTIEAESVKAARARLKRIGFDLLGYFKSMGAVAH 190
 Query: 177 LLGGEERVNLADMLRGENHL-PFSYKDLTSLGQSTKHFIAPT YLSFKHKNHIELDDRLL 235
 ++ G ER+N+L + + + F +K L SG STK FIAP+ L F + + +
 10 Sbjct: 191 VMDGWERLNLHGVYHPDGEIFNFDWKWLAPSGSLTKDFIAPSSLCFGNAKTFGMGGKYG 250
 Query: 236 QIVYVRDYGMEIGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQ 295
 + +++ EL D + D + ++ V+++LH + +++ + ++ K T +++ KI EQ
 15 Sbjct: 251 AVSFLQILSPELSDDMLADFLNTESGVLVNLHVQAIEQTKAIKTKRKITDL DAMKIAEQ 310
 Query: 296 QKMATGIYLEKVGHVLENNIDEAEALLQMTQTGDKLFDITVFLIGVLADTEDQLKQSLD 355
 +K R+G ++ + L ++A+ LL + ++LF FL+ +ADT+ +L +
 Sbjct: 311 KKAVERSGYDMDILPSDLATYGEDAKKLLTKLQTRNERLFLQTLFVLNVADTKQKLNNDVF 370
 Query: 356 IIKQVAGSNDMIIDNLT YMQEAAFNSLLPFGKNYLEGVSRSLTSNIAVNAPWTSVDIHD 415
 VA ++ + L Y QE S LP G N ++ + RSL TS++AV P+ + ++
 20 Sbjct: 371 QAAGVAQKHNCPLVRLDYQQEQGLASSPLGVNQIK-IQRSLTSSVAVFVPFVTQELFQ 429
 Query: 416 KGGK-FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDE 474
 G +YGIN S N+I +DR + P+ L LGT G+GK M+ K EI+S L D +
 25 Sbjct: 430 GGAAMYYGINAKSRNMIMLDRQARCPNALKLGTGSGKSMCKSEIVSVFLTPD---D 486
 Query: 475 IIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELS-DENMDEDPVKVSEFLLSWIG 533
 I I DPE EY + + G+ I ++P S F+N L+++ + + D++P+ +KS+F+LS+
 30 Sbjct: 487 IPISDPAEYYPVLRHLGQVIRLSPTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCE 546
 Query: 534 KLLDRK--MDGREKSLIDRVTRLTYKHF-----DTPSLVEWVFLVLSQQPEQEA KDIAL 584
 ++ K ++ EK++IDR R+ Y+ + + P L + L Q EA +A
 35 Sbjct: 547 LVMGGKNGLEAIEKTVIDRAVRVIYRPLYADPRPENMPILSDLHKALLDQHVPEADRV AQ 606
 Query: 585 DMELYVEGSLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVI FDQIWNRVVKNQKLG 644
 ++LYV GSL++F+HRTN+ + + +++K+LG +LK++ ++++ DQIW RV N+ G
 Sbjct: 607 ALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKKLGMLIVQDQIWGRVTANRSQG 666
 Query: 645 KKTWYIFDEMQLLLLDKYASDFFFKLSRVRKYGAIPTGITQNVETLLLDANGRR IANS 704
 K TW + DE LLL ++ + + ++W R RK+G IPTG TQNV+ LL I+ NS
 40 Sbjct: 667 KATWYFADEFHLLKKEEQTAAYSAEIKRFRKGGIPTGATQNVKDLLSSPEIENILENS 726
 Query: 705 EFMILLKQAKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLF 764
 +F+ LL QA DR+ L L LS E +KY+ N E G GL+ + V+PF N IP +T+L+
 45 Sbjct: 727 DFITLLNQASGDRKILAEERLNLSTEQQKYIDNSEPGEGLLIFENVVL PFTNPIPHNTQLY 786
 Query: 765 DIMST 769
 IM+T
 50 Sbjct: 787 KIMTT 791

SEQ ID 8926 (GBS75) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 11; MW 89.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 6; MW 114.7kDa).

55 GBS75-GST was purified as shown in Figure 197, lane 8.

GBS329 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 8; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 174 (lane 2; MW 114kDa).

GBS329-GST was purified as shown in Figure 220, lanes 9 & 10.

-2140-

Example 1891

A DNA sequence (GBSx1999) was identified in *S.agalactiae* <SEQ ID 5875> which encodes the amino acid sequence <SEQ ID 5876>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 33
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2442(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1892

A DNA sequence (GBSx2000) was identified in *S.agalactiae* <SEQ ID 5877> which encodes the amino acid sequence <SEQ ID 5878>. This protein is predicted to be DNA-directed RNA polymerase ii largest subunit. Analysis of this protein sequence reveals the following:

```

20   Possible site: 21
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.4393(Affirmative) < succ>
25   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1893

A DNA sequence (GBSx2001) was identified in *S.agalactiae* <SEQ ID 5879> which encodes the amino acid sequence <SEQ ID 5880>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 13
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -9.92    Transmembrane  256 - 272 ( 250 - 277)
    INTEGRAL    Likelihood = -8.28    Transmembrane  216 - 232 ( 213 - 244)
    INTEGRAL    Likelihood = -8.12    Transmembrane  151 - 167 ( 148 - 191)
40   INTEGRAL    Likelihood = -7.27    Transmembrane   57 -  73 (  54 -  80)
    INTEGRAL    Likelihood = -6.74    Transmembrane   93 - 109 (  88 - 111)
    INTEGRAL    Likelihood = -3.50    Transmembrane  172 - 188 ( 168 - 191)
    INTEGRAL    Likelihood = -2.76    Transmembrane  113 - 129 ( 110 - 130)

45   ----- Final Results -----
    bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-2141-

>GP:AAG38039 GB:AF295925 Orf23 [Streptococcus pneumoniae]
Identities = 71/86 (82%), Positives = 83/86 (95%)

Query: 37 VKSLADFNPTVWSYMTAITKGIMQPLGVAILAVVLVLEFSKMAKKIANS GGAMTFEAIAP 96
+KSL+ +NPTVW+YM++ITK +MQPLGVAIL+VVL+LEFSKMAKKIANS GGAMTFEA+AP
Sbjct: 1 MKSLSSYNPTVWTYMSITKSVMQPLGVAILS VVLILEFSKMAKKIANS GGAMTFEALAP 60

Query: 97 MIVSYIMVAVVITNTTTVIVEAIIAIA 122
M++SYIMVAVVITNTTTVIVEAII IA
Sbjct: 61 MLISYIMVAVVITNTTTVIVEAIIIGIA 86

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 1894

A DNA sequence (GBSx2002) was identified in *S.agalactiae* <SEQ ID 5881> which encodes the amino acid sequence <SEQ ID 5882>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

| | | | |
|----------|--------------------|---------------|--------------------|
| INTEGRAL | Likelihood = -7.54 | Transmembrane | 32 - 48 (25 - 52) |
| INTEGRAL | Likelihood = -4.09 | Transmembrane | 63 - 79 (62 - 80) |

----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.4015(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

A related GBS nucleic acid sequence <SEQ ID 9933> which encodes amino acid sequence <SEQ ID 9934> was also identified. A related GBS nucleic acid sequence <SEQ ID 10777> which encodes amino acid sequence <SEQ ID 10778> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1895

A DNA sequence (GBSx2003) was identified in *S.agalactiae* <SEQ ID 5883> which encodes the amino acid sequence <SEQ ID 5884>. This protein is predicted to be TrsK-like protein (traK). Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have a cleavable N-term signal seq.

| | | | |
|----------|--------------------|---------------|--------------------|
| INTEGRAL | Likelihood = -7.38 | Transmembrane | 66 - 82 (62 - 85) |
|----------|--------------------|---------------|--------------------|

----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.3951(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38037 GB:AF295925 Orf21 [Streptococcus pneumoniae]

Identities = 343/457 (75%), Positives = 385/457 (84%), Gaps = 24/457 (5%)

-2142-

Query: 142 LIVIGSGAGKTFRFVKPNLIQLNCSNIVVDPKDHAEKTGKLFLENGYQVKVLDLVNMT 201
 + VIGSG+GKTFRFVKPNLIQ+N SNIVVDPKDHAEKTGKLFLE+GYQVKVLDLVNM
 Sbjct: 1 MAVIGSGSGKTFRFVKPNLIQMSSNIVVDPKDHAEKTGKLFLEHGYQVKVLDLVNMK 60

Query: 202 NSDGFNPFRYVETENDLNRLMTVYFNNTKNGSRSDPFWEASMTLVRAIASYLVDFFYNP 261
 NSDGFNPFRY+ETENDLNRLM VYFNNTKG+GSRSDPFWEASMTLVRA+ASYLVDFFYNP
 Sbjct: 61 NSDGFNPFRYIETENDLNRLMAVYFNNTKNGSRSDPFWEASMTLVRAIASYLVDFFYNP 120

Query: 262 PGS-----SKQEAEARRKGRYPAFSEIGKLIKLSKGDNDQDKS 300
 P + K+E E R+KRGR F E + + + KS
 Sbjct: 121 PKTREQLIEESRLSQKEYQNLLKQKKEVEERKKRGLSKFCESQNSLNTYPRVKTR-KS 179

Query: 301 ILEVLFEDYAKKYGHENFTMRNWADFQNYKDKTLDSEVIAVTTAKFALFNIQSVIDLTQRD 360
 +LE+LFE+YAKKYG ENFTMRNWADFQNYKDKTLDSEVIAVTTAKFALFNIQSV+DLT+RD
 Sbjct: 180 VLEILFENYAKKYGTENFTMRNWADFQNYKDKTLDSEVIAVTTAKFALFNIQSVMDLTQRD 239

Query: 361 TMDLKTWGTQKTMVYLVPDNDTTFRFLSAL-FFSTVFSTLTROADV-DFKGQLPIHVR 418
 T+D+KTWG +K+MVYLVPDND+TFRFLSAL FF+ F T + + + +LP+HVR
 Sbjct: 240 TLDKMTWQEKSMVYLVPDNDSTFRFLSALLFFNPYFQTPNKTSQILMLRVRLPLHVRV 299

Query: 419 YLDEFANVGEIPDFAEQTSTVRSRNMVLPILQNIQGLYKEKEAWKTILGNCDLLY 478
 YLDEFAN+GEIPDFAEQTSTVRSRNMVLPILQNIQGLYKEKEAWKTILGNCDL+Y
 Sbjct: 300 YLDEFANIGEIPDFAEQTSTVRSRNMVLPILQNIQGLYKEKEAWKTILGNCDLVY 359

Query: 479 LGGNDEETFFKMSGLLGKQTVDRSTSRSGQTGSSSTSHQKIARDLMTADEVGTMRDE 538
 LGGNDE+TFKMSGLLGKQTVDR+TSRSGQTGS S SHQKIARDLMT DEVG MKR E
 Sbjct: 360 LGGNDEDTFFKMSGLLGKQTVDRNTSRSGQTGSGSLSHQKIARDLMTPEVGNMKRHE 419

Query: 539 CLVRIAGVPVFRTKKYFPLKHKHKLADKETDDRWW 575
 CLVRIA +PVF++KKY KH +WK LA++ETD+R W
 Sbjct: 420 CLVRIANMPVFKSKYNSTKHPNWKYLANQETDERRW 456

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8929> and protein <SEQ ID 8930> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5
 McG: Discrim Score: 5.53
 GvH: Signal Score (-7.5): -0.78
 Possible site: 34
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -7.38 threshold: 0.0
 INTEGRAL Likelihood = -7.38 Transmembrane 66 - 82 (62 - 85)
 PERIPHERAL Likelihood = 1.75 338
 modified ALOM score: 1.98

*** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

33.9/50.9% over 419aa

Lactococcus lactis

GP|3582206| trsK protein (traK) Insert characterized

PIR|T43089|T43089 transfer complex protein TrsK - plasmid pMRC01 Insert characterized

ORF00383(715 - 2004 of 2415)

-2143-

```

GP|3582206|gb|AAC56002.1||AE001272(23 - 442 of 530) trsK protein (traK) {Lactococcus
lactis|PIR|T43089|T43089 transfer complex prote
in TrsK - Lactococcus lactis plasmid pMRC01
%Match = 10.1
%Identity = 33.8 %Similarity = 50.8
Matches = 141 Mismatches = 193 Conservative Sub.s = 71

```

519 549 579 609 639 669 699 729
SFLAFILGLVLMMTLVLYLVSTGQKVYREGEEYGSA RFGTSKEKRN FYSKNPFNDTILARDVRLTLEKKKPQFDNRNKL I
| : | ::
MNGTILGLVDNKIIYQDNTTKPNRNV M
10 20

759 789 816 846 876 906 936 966
VIGGSGAGKTRFRVVKPNLIQLNCSNIVV-DPKDHLAEKTKGLFLENGYQVKVLDLVNMTNSDGFNPFYRVETENDLNRLM
| | | | | : | | | | : | | | | | | | : | | : | : : | | : | : | : | : | : | :
VIGGSGSYKTSQVWITNLFNETKSNIVVTDPKGELYEKTAGIKLAQGYEVHVVNFANMAHSDRYNPFYIERIDQAESVA
 40 50 60 70 80 90 100

996 1026 1056 1086 1116 1146 1176 1194
TVYFNNKTGNGSRSDPFWDEASMTLVRAIASYLVDFYNPPGSSKQEQEARKRGRYPAFSEIGKLIKLLSKGD---NQD
| : | : | : : : : : : : :
TKIVQSENAEGKK--DVWFSTQRQLKALILFVM-----KERSPEQRNLAGVINVLQTFDSEPINKD

K-SILEVLFDYAKKYGHENFTMRNWADFQNYKDKTLDSVIAVTAKFALFNISVIDLTQRDTMDLKTWGTQKTMVLYV
:
| : | : || | | | | | | | : | | | : | : | : | : | | | : | : | : | :
ENSDDLNLFF--LALKITHPARIAYELG-FKKAGKDMKASISSLATISKFTDEEVSNFTSISDFHQLDIGRKKIVLYVI

180 190 200 210 220 230 240

1461 1491 1521 1551 1581 1611 1641 1671
IPDNDTTFRFLSALFFSTVFSLTLTQADVDFKGLPIHVSRYLDEFANVGEPIDFAEQTSTVRSRNMSLVPILQNIQLQ
|| | |: : ||| |: | |: : ||| |: |: |: |: | |: : | |: ||
IPVMDNTYESFINLFFSQMFDELYKLASSN-GAKLPQEVDFILDEFVNGLGKFPKYEEFLATCRGYGIGVTTICQTLTQLQ
260 270 280 290 300 310 320

1701 1731 1761 1791 1809 1839 1869 1899
GLYKEKEAWKTLGNCDLLYLGGNDEETPKFMSGLLGKQTVDVDR---STSRSGQTGSSSTSHQKIARDLMTADEVGT
|| || ::||| : | ::| | : | ||| || | ||| | : | | : : | ||| ||:
SLY-GKEKAESILGNHAVKICLNASNEATAKYFSELLGKSTVKVETGSESTSHSKETSKSDSYSTSRQLMTPDEIIR
340 350 360 370 380 390 400

1929 1956 1974 2004 2034 2064 2094 2124
MKRDECLVRIAGV-PVPRTK---KYFPLKHKHWKLLADKETDDRWWNYHINPLAKEEELDLSDYQIRDLSLETSLH**K
| : | : | : || | || | : | : : : :
MPDTQSLLIFTNQPKPIKATKAFOFKLFPDADSQVKLEQNKYVGITSKSQLEKYNDLSVKWEEKLSQLKNITVTETEEEDKL

420 430 440 450 460 470 480

SEQ ID 5884 (GBS11d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 6; MW 61kDa) and in Figure 182 (lane 10; MW 61kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 5; MW 91.5kDa).

Example 1896

A DNA sequence (GBSx2004) was identified in *S.agalactiae* <SEQ ID 5885> which encodes the amino acid sequence <SEQ ID 5886>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.4192 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-2144-

A related GBS nucleic acid sequence <SEQ ID 9935> which encodes amino acid sequence <SEQ ID 9936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1897

10 A DNA sequence (GBSx2005) was identified in *S.agalactiae* <SEQ ID 5887> which encodes the amino acid sequence <SEQ ID 5888>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1898

25 A DNA sequence (GBSx2006) was identified in *S.agalactiae* <SEQ ID 5889> which encodes the amino acid sequence <SEQ ID 5890>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have an uncleavable N-term signal seq

30 INTEGRAL Likelihood = -10.03 Transmembrane 68 - 84 (64 - 90)
 INTEGRAL Likelihood = -7.06 Transmembrane 33 - 49 (27 - 50)
 INTEGRAL Likelihood = -5.73 Transmembrane 106 - 122 (105 - 123)
 INTEGRAL Likelihood = -4.46 Transmembrane 6 - 22 (3 - 24)
 INTEGRAL Likelihood = -2.13 Transmembrane 154 - 170 (154 - 170)
 INTEGRAL Likelihood = -0.53 Transmembrane 180 - 196 (180 - 196)

35 ----- Final Results -----

 bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9937> which encodes amino acid sequence <SEQ ID 9938> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA11325 GB:D78257 ORF8 [Enterococcus faecalis]

Identities = 35/102 (34%), Positives = 57/102 (55%), Gaps = 4/102 (3%)

45

Query: 90 TRNQAVLVQVGKQVPPIIFLLFL-VNASILEEIVYRQLLWEKLTFF--PFEQIGVTSFLFV 146
 T N + L+++ V P++ +L L + A I+EEIV+R + L I ++SFLF
 Sbjct: 7 TANDSTLIKLFSGVSPVLVLLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILISSFLFG 66

Query: 147 LSHGPNQLGSWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWN 187
+ HGP S+ +Y +G+ L+V .KT D +I++H L N

Subject: 67 IIHGPTNFIISFGMYFFMGIIILSVSYKYTKDLRVSIISIHFLNN 108

A related GBS gene <SEQ ID 8931> and protein <SEQ ID 8932> were also identified. Analysis of this protein sequence reveals the following:

```

bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

774 804 831 861 891 921 951 981
HGPNQLGSLWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWNSLAYVVTFL*YQNQECFRIMEAPYV**GIEKRGGHYVI*T
||| : |: :| :|: |:| ||| :|: |:| |
HGPTNFISFGMYFFMGIILSVSYKTKDLRVSISIHFLNNLFPATAIAYGLI
80 90 100 110 120

Possible site: 23

-2146-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2490(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9939> which encodes amino acid sequence <SEQ ID 9940> was also identified.

10 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1900

15 A DNA sequence (GBSx2008) was identified in *S.agalactiae* <SEQ ID 5893> which encodes the amino acid sequence <SEQ ID 5894>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.5298(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98423 GB:L29323 unknown [Streptococcus pneumoniae]
 Identities = 68/126 (53%), Positives = 88/126 (68%)

30 Query: 1 MNLHKKSIILDCLEERIHQAETNQLLQKILSLPNFDCDFEVTFFEDDYHKEMNDPLFYE 60
 M L+K+SILDC ELE +H AE QL ++I +PN+ C+FEVTF DDYHK+ N PLFYE
 Sbjct: 1 MKALNKESILDCDELETSLHDAEIKQLDEQIFLMPNYPCEFEVTFLLDDYHKKHNYPLFYE 60

 Query: 61 SNLHQISDFMETRDIKNGVDTLLTKDNHLAFRAFGENYSARGKEGILTTLVTVKCFGEGR 120
 S L I +F+E++DIKNG D + +L F +G+ Y A GKEGILT VTVK F E +
 35 Sbjct: 61 SYLQNIIMEFLSQDIKNGADAFVDDHQNLFVFLYQGYRAEGKEGILTTQVTVKAFDEDK 120

 Query: 121 MPIDMS 126

 PI+ +

40 Sbjct: 121 KPINF 126

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1901

45 A DNA sequence (GBSx2009) was identified in *S.agalactiae* <SEQ ID 5895> which encodes the amino acid sequence <SEQ ID 5896>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

-2147-

bacterial cytoplasm --- Certainty=0.1209(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98421 GB:L29323 methyl transferase [Streptococcus pneumoniae]
 Identities = 323/449 (71%), Positives = 389/449 (85%), Gaps = 3/449 (0%)

Query: 1 MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
 M+F+DLF+GIGGFRLGMES GH+C+GFCEIDKFAR SYK++F TEGEIE+HDI++V+D +
 Sbjct: 1 MRFDLDFSGIGGFRLGMESVGHCEIGFCEIDKFARESYKSIFQTEGEIEFHDIRDVSDD 60

Query: 61 FRQFRGQVDIIICGGFPCQAFSLAGRRIGFEDTRGTLFFEIARAQKIQPRFLFLENVKGL 120
 F++ RG+VD+ICGGFPCQAFS+AGRRIGFEDTRGTLFFEIARAQKIQPRFLFLENVKGL
 Sbjct: 61 FKLRGKVDVICGGFPCQAFSIAAGRRIGFEDTRGTLFFEIARAQKIQPRFLFLENVKGL 120

Query: 121 LNHDEGRFTATILSTLDELGYDVEWQVLNSKDFQVQNRRERVFIIGHSRYSRFFIFPLR 180
 LNHD+GRTF TIL+TLDELG+DVEWQ+LNSKDF VQNRRERVFIIGHSR+ +R FP R
 Sbjct: 121 LNHDKGRFTTILSTLDELGYDVEWQVLNSKDFGVPQNRRERVFIIGHSRKRGTRLGFPFR 180

Query: 181 RED---SPAHLERLGNINPSKHGLNGEVYLTSLGLAPTLTRGKGEGAKIAIPVLTPDRLEK 237
 RE +P L+ LGN+NPSK G++G+VY + GLAPTL RGKGEG KIAIP +TPDRL+K
 Sbjct: 181 REGQATNPETLKILGNLNPSSKSGMSGVYSEGLAPTLVRGKGEGFKIAIPCMTPDRLDK 240

Query: 238 RQHGRFRKDNQDPMFTLTSDKHGVVAGNLPTSFDQTGRVFDISGLSPTLTMTQGGDKV 297
 RQ+GRRFKDNQ+PMFTL +QD+HG+VV G+LPTSF +TGRV+ GLSPTLTMTQGGDK+
 Sbjct: 241 RQNGRRFKDNQEPMTLTNTQDRHGIVVVDLPTSFKETGRVYSGELSPTLTMTQGGDKI 300

Query: 298 PKILLREELPFLKIKEATKTGYAKATLGDSVNLAYPDSTKRRGRVVGKISNTLTSDNMG 357
 PKIL+ E + FLK++EATK GYA+A +GDS+NL P S RRGRVVGKI+NTLTTS MG
 Sbjct: 301 PKILIPEPIQFLKVREATKKGYAQAEGDSINLERPSSQHRRGRVVGKIANLTTSQMG 360

Query: 358 VVVALEYRQDKWYEVGTIVLEGKLYRLRIRRLTPRECRLQGFPDWAYERAESVSSKSQ 417
 VVVA+ E + Y+V G++++G+ YRLRIRR+TP+ECFRLQGFPDWA+E A VSS SQ
 Sbjct: 361 VVVASYEGEDKQVYQVAGVLIDGQFYRLRIRRLTPKECFRLQGFPDWAFAARKVSSNSQ 420

Query: 418 LYKQAGNSVTVTVIEAIAREFRRTEEBEK 446
 LYKQAGNSVTV VI AIA++ + EE+++
 Sbjct: 421 LYKQAGNSVTVPVIAIAAKKLKEVEEKDE 449

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2435> which encodes the amino acid sequence <SEQ ID 2436>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1725(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/75 (80%), Positives = 69/75 (92%)

Query: 1 MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
 MKFLDLFAGIGGFRLG+ +Q H+C+GFCEIDKFAR SYKA++ TEGEIE+HDI++VTD D
 Sbjct: 4 MKFLDLFAGIGGFRLGLINQCHECIGFCEIDKFARQSYKAIYETEGEIEFHDIRQVTDQD 63

Query: 61 FRQFRGQVDIIICGGF 75
 FRQ RGQVDIIICGGF
 Sbjct: 64 FRQLRGQVDIIICGGF 78

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2148-

Example 1902

A DNA sequence (GBSx2010) was identified in *S.agalactiae* <SEQ ID 5897> which encodes the amino acid sequence <SEQ ID 5898>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 16
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood = -9.71    Transmembrane    8 - 24 ( 3 - 30)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9941> which encodes amino acid sequence <SEQ ID 9942> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5899> which encodes the amino acid sequence <SEQ ID 5900>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 28
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL    Likelihood = -1.81    Transmembrane    20 - 36 ( 19 - 36)

      ----- Final Results -----
25          bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 16/33 (48%), Positives = 26/33 (78%)

30      Query: 1  MNKMIWWILGGIYLISIIILIVEIIRAPEMDDH 33
          ++KM WW+L G++ + I LI+E+I APEM+D+
      Sbjct: 12  VSKMFWWLLGLGVWGLRTIWLIIIEVITAPEMEDY 44

```

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1903

A DNA sequence (GBSx2011) was identified in *S.agalactiae* <SEQ ID 5901> which encodes the amino acid sequence <SEQ ID 5902>. This protein is predicted to be ifn-response binding factor 1 (irebf-1). Analysis of this protein sequence reveals the following:

```

40      Possible site: 53
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
45          bacterial cytoplasm --- Certainty=0.4771(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50      >GP:AAD41248 GB:AF106927 unknown [Streptococcus suis]
          Identities = 258/272 (94%), Positives = 266/272 (96%)

      Query: 1  MKRITANQYQTSERYKLPKILFESERYKDMKLEVKVAVAVLKDRLELSLSKGWIDEDGA 60
          MKRITANQYQTSERYKLPKILFESERYKDMKLEVKVAVAVLKDRLELSLSKGWIDEDGA

```

-2149-

Sbjct: 1 MKRITANQYQTSERYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA 60

Query: 61 IYLIYSNSNLMLLGCSSKLLSIKKTLEFYGLIDEVQSSSERGRMANKIYLGELEHEP 120
IYLIYSNSNLMLLGCSSKLLSIKKTLEFYGLIDEVQSSSE+GRMANKIYLGELEHE

5 Sbjct: 61 IYLIYSNSNLMLLGCSSKLLSIKKTLEFYGLIDEVQSSSEKGRMANKIYLGELEHET 120

Query: 121 TPVLHTDGASVKKTLGESQKRTGTPVLYSAPSETEGSETKYSETEGSDLVMKDEEERQLVD 180
TPVLHTDGASVKKTLG SQRKTGPVL SAPSETEGSETKYSET+GSD +++DEEERQ VD

10 Sbjct: 121 TPVLHTDGASVKKTLGGSQKRTGTPVLNSAPSETEGSETKYSETKGSDFLIEDEEERQQVD 180

Query: 181 EKKEENFTSKVDGVTYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMR 240
EK+EENFTSKVDGVT+YDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALE MRF

Sbjct: 181 EKQEENFTSKVDGVTRYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEHMR 240

15 Query: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGG 272
ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGG+

Sbjct: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGGD 272

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5903> which encodes the amino acid
20 sequence <SEQ ID 5904>. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.5248(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 84/122 (68%), Positives = 99/122 (80%), Gaps = 2/122 (1%)

Query: 145 VLYSAPSETEGSETKYSETEGSDLVKDEEERQLVD--EKKEENFTSKVDGVTKYDRDYI 202
VL SAPSETE SET+ SET+ S+LV++DEEER+ +K E +FT +VD VTKYD+DYI

35 Sbjct: 1 VLNSAPSETKSETEGSETKESNLVIEDEEERKECTSVKKTEGHFTRQVDQVTKYDKDYI 60

Query: 203 WGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRFARSAEVIAEYVFNGVLSEWTKQ 262
W LVH QLR+ GLSQ+ASD M YF +RY YALE +RFAR+AE IAEYVFNGVLSEWTKQ

Sbjct: 61 WSLVHSQLRREGGLSQAASDLVMSYFEERYAYALEHIRFARTAEIAEYVFNGVLSEWTKQ 120

40 Query: 263 LR 264
LR

Sbjct: 121 LR 122

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
45 vaccines or diagnostics.

Example 1904

A DNA sequence (GBSx2012) was identified in *S.agalactiae* <SEQ ID 5905> which encodes the amino
acid sequence <SEQ ID 5906>. Analysis of this protein sequence reveals the following:

Possible site: 17
50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.4191(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9375> which encodes amino acid sequence <SEQ ID 9376>
was also identified.

-2150-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1905

A DNA sequence (GBSx2013) was identified in *S.agalactiae* <SEQ ID 5907> which encodes the amino acid sequence <SEQ ID 5908>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3723 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 1906

A DNA sequence (GBSx2014) was identified in *S.agalactiae* <SEQ ID 5909> which encodes the amino acid sequence <SEQ ID 5910>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3053 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1907

A DNA sequence (GBSx2015) was identified in *S.agalactiae* <SEQ ID 5911> which encodes the amino acid sequence <SEQ ID 5912>. This protein is predicted to be 50S ribosomal protein L7/112 (rpL). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1034 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

-2151-

A related GBS nucleic acid sequence <SEQ ID 9943> which encodes amino acid sequence <SEQ ID 9944> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAB11881 GB:Z99104 ribosomal protein L12 (BL9) [Bacillus subtilis]
    Identities = 83/123 (67%), Positives = 95/123 (76%), Gaps = 2/123 (1%)

    Query: 6  MALNIENIIAEIKEATILELNDLVKAIEEEFGVTAAAPVAAA--AAGGEAAAADKDSFQVE 63
              MALNIE IIA +KEAT+LELNDLVKAIEEEFGVTAAAPVA A AA G AA + FD+
10  Sbjct: 1  MALNIEEIIASVKEATVLELNDLVKAIEEEFGVTAAAPVAVAGGAAAGGAAEEQSEFDLI 60

    Query: 64  LTAAGDKKVGVIKVVREITGEGLEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAGASV 123
              L AG +K+ VIKVVREITG GLKEAK +VDN P +KEG ++ EA E+K KLE GASV
15  Sbjct: 61  LAGAGSQKIKVIKVVREITGLGLKEAKELVDNTPKPLKEGIAKEEAEELKAKLEEVGASV 120

    Query: 124 TLK 126
              +K
    Sbjct: 121 EVK 123
  
```

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5913> which encodes the amino acid sequence <SEQ ID 5914>. Analysis of this protein sequence reveals the following:

```

    Possible site: 51
    >>> Seems to have no N-terminal signal sequence
  
```

```

25  ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1164(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

30 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 104/126 (82%), Positives = 113/126 (89%)
  
```

```

    Query: 1  MEEITMALNIENIIAEIKEATILELNDLVKAIEEEFGVTAAAPVAAAAAGGEAAAADKDSF 60
              +EETMALNIENIIAEIKEA+ILELNDLVKAIEEEFGVTAAAPVAAAAAGG AAKDSF
35  Sbjct: 1  LEEITMALNIENIIAEIKEASILELNDLVKAIEEEFGVTAAAPVAAAAAGGAREAAKDSF 60

    Query: 61  DVELTAAGDKKVGVIKVVREITGEGLEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAG 120
              DVELT+AGDKKVGVIK VREITG GLKEAK +VD AP+ +KEG + AEA EIK KLE AG
40  Sbjct: 61  DVELTSAGDKKVGVIKAVREITGLGLKEAKGLVDGAPANVKEGVAAAEAEIKAKLEEAG 120

    Query: 121 ASVTLK 126
              A++TLK
    Sbjct: 121 ATTTLK 126
  
```

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1908

A DNA sequence (GBSx2017) was identified in *S.agalactiae* <SEQ ID 5915> which encodes the amino acid sequence <SEQ ID 5916>. This protein is predicted to be ribosomal protein L10 (rplJ). Analysis of this protein sequence reveals the following:

```

    Possible site: 37
    >>> Seems to have no N-terminal signal sequence
  
```

```

55  ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1251(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  
```

-2152-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11880 GB:Z99104 ribosomal protein L10 (BL5) [Bacillus subtilis]
 Identities = 96/164 (58%), Positives = 125/164 (75%), Gaps = 1/164 (0%)

Query: 14 MSEAIIAKKAQVELIAEKMKAASIVVDSRGLTVEQDTNLRSLRESDFEFKVIKNSI 73
 MS AI KK VE IA K+K + S ++VD RGL V + T LR+ LRE++VE KV KN++

Sbjct: 1 MSSAIETKKVV-VEEIASKLKESKSTIIVDYRGLNVSEVTELRKQLREANVESKVYKNTM 59

Query: 74 LTRAAEKAGLEDLKEFLVGPSSAVAFSNEVDVIAPAKVISDFAKDAEAELEIKGGSVDGKFTS 133
 RA E+A L L + GP+A+AFS EDV+APAKV++DFAK+ EALEIK G ++GK ++

Sbjct: 60 TRRAVEQAELNGLNDFLTGPNIAIAFSTEDVVAPAKVLNDFAKNHEALEIKAGVIEGKVST 119

Query: 134 VEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEDKEE 177
 VEE+ ALA+LP +EG+LSMLLSVL+APVRN+A A KAVAE+ EE

Sbjct: 120 VEEVKALAEPLPPREGLLSMLLSVLKAPVRNLALAAKAVAEQKEE 163

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5917> which encodes the amino acid sequence <SEQ ID 5918>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -5.47 Transmembrane 7 - 23 (5 - 24)

----- Final Results -----

bacterial membrane --- Certainty=0.3187(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 149/176 (84%), Positives = 162/176 (91%)

Query: 4 SQKIKTEVKLMSEAIIAKKAQVELIAEKMKAASIVVDSRGLTVEQDTNLRSLRES 63

S KIKTEVKLMSEAIIAKKAQVELIAEKMKAASIV+VDSRGLTV+QDT LRRSLRES

Sbjct: 23 SPKIKTEVKLMSEAIIAKKAQVELIAEKMKAASIVIVDSRGLTVQDTVLRRSLRES 82

Query: 64 VEFKVIKNSILTRAAEKAGLEDLKEFLVGPSSAVAFSNEVDVIAPAKVISDFAKDAEAELEIK 123

VEFKVIKNSILTRAAEKAGL++LK++FVGPSAVAFSNEVDVIAPAKVI+DF K A+ALEIK

Sbjct: 83 VEFKVIKNSILTRAAEKAGLEDLKDVFVGPSAVAFSNEVDVIAPAKVINDFTKTADALEIK 142

Query: 124 GGSVDGKFTSVEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEDKEEVA 179

GG+++G +S EEI ALA LPN+EGMLSMLLSVLQAPVRNVAYAVKAVAE E A

Sbjct: 143 GGAIEGAVSSKEEIQALATLPNREGMLSMLLSVLQAPVRNVAYAVKAVAENKEGAA 198

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1909

A DNA sequence (GBSx2018) was identified in *S.agalactiae* <SEQ ID 5919> which encodes the amino acid sequence <SEQ ID 5920>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.22 Transmembrane 125 - 141 (106 - 143)

INTEGRAL Likelihood = -1.91 Transmembrane 108 - 124 (106 - 124)

----- Final Results -----

bacterial membrane --- Certainty=0.3888(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2153-

A related GBS nucleic acid sequence <SEQ ID 10931> which encodes amino acid sequence <SEQ ID 10932> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1910

- 10 A DNA sequence (GBSx2019) was identified in *S.agalactiae* <SEQ ID 5921> which encodes the amino acid sequence <SEQ ID 5922>. This protein is predicted to be Clp-like ATP-dependent protease binding subunit (clpC). Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 15 bacterial cytoplasm --- Certainty=0.3483(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:AAA68910 GB:L34677 Clp-like ATP-dependent protease binding
 subunit [Bos taurus]
 Identities = 437/589 (74%), Positives = 514/589 (87%), Gaps = 5/589 (0%)
- 25 Query: 10 DPFGN-MDDIFNSIMGNMGGYNSENKRYLINGREVTPEEFSQYRQTGKLPQGELNNQNTTP 68
 DPF N MDD+FN LMG M G NSEN+RYLINGREVTPEE++ +RQTGKLPG Q
 Sbjct: 2 DPFNNDMDLFLNQLMGGMNGVNSENRRYLINGREVTPEEYAAFRQTGKLPGVTDPTQ-AK 60
- 30 Query: 69 TNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAELARRTKNNPVLVGDAVG 128
 T Q DS+L KLG NLT +A++ LDPVIGRNKEIQETAEL+RRTKNNPVLVGDAVG
 Sbjct: 61 TKQPQPSMLAKLGRNLTQEAKEGKLDPVIGRNKEIQETAELSRRTKNNPVLVGDAVG 120
- 35 Query: 129 KTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIIEVKETGN 188
 KTAV+EGLAQAI+ GDVPAAIKNK+IISIDISSLEAGTQYRGSFEEN+Q +I EVK+ GN
 Sbjct: 121 KTAVVEGLAQAIIVAGDVPAAIKNKQIISIDISSLEAGTQYRGSFEENMQKLIDEVKKDGN 180
- 40 Query: 189 IILFFDEIHQILGAGSTGGDSGSKGLADILKPALSRGELTVIGATTQDEYRNTILKNAAL 248
 +ILFFDEIHQI+GAG+ G SGSKG+ADILKPALSRGE+T+IGATTQDEYRNTILK+AAAL
 Sbjct: 181 VILFFDEIHQIIGAGNAGDASGSKGMADILKPALSRGEVTLIGATTQDEYRNTILKDAAL 240
- 45 Query: 249 ARRFNEVKVNAPSAQDTFNIILMGIRNLYEQHHNVLPDSVLKAAVDLSIQYIPQRSPLDK 308
 +RRFN+V VNAPS +DTF IL G+R LYE+HHNV LPD VLKAA+D S+QYIPQRSPLDK
 Sbjct: 241 SRRFNQVTVNAPSKEDTFKILQGLRKLYEKHHNVSLPDEVLKAAIDYSVQYIPQRSPLDK 300
- 50 Query: 309 AIDLIDMTAAHLAAQHPVTDLKSLEKEIAEQDKQEKAVNTEDFEEALKVKTRIEELQNO 368
 AIDLID+TAHLA++HPV D K++E+EI + KQ++AV ED++ A + K ++ +LQ+Q
 Sbjct: 301 AIDLIDVTAHLASKHPVKDAKTEEEIKKTEAKQQEAVEKEDYQAAQEAQKQVAKLQDQ 360
- 55 Query: 369 IDNHTGQKVTATINDIAMSIERLTGVPVSNMGASDIERLKEGNRLKGKVGIGNDAVEA 428
 + +H+E ++V AT +D+A ++ER+TG+PVS MGASDIERL L RL+GKVGQ +AVEA
 Sbjct: 361 LKDHSESESVVATPSDVAAVERMTGIPVSKMGASDIERLKGATRLGKVGIGQVEAVEA 420
- Query: 429 VARAIRNRAGFDDGNRPIGSFLFVGPTGVGKTELAQQLAFDMFGSKDAIVRLDMSEYND 488
 V+RAIRNRAGFD+GNRPIGSFLFVGPTGVGKTELAQQLA DMFGS + I+RLDMSEY D
 Sbjct: 421 VSRAIRNRAGFDEGNRPIGSFLFVGPTGVGKTELAQQLALDMFGSTNDIIRLDMSEYTD 480
- Query: 489 RTAVSKLIGATAGYVGYYDDNSNTLTERIRRNYPYSIVLLDEIEKADPQVITLLQLVDDGR 548
 RTAVSKLIG TAGYVGYYDDNSNTLTE++RR+PYSIVLLDEIEKA+PQVITLLQLVDDGR

-2154-

Sbjct: 481 RTAVSKLIGTTAGYVGYDDNSNTLTERKRRHPYSIVLLDEIEKANPQVITLLQLVDDGR 540

Query: 549 LTDGQGNTINFKNTVIIATSNAGFGNEAFTGDSKDLKIMERISPYFRP 597

LTDGQGNT++FKNT+IIATSNAGF ++A G+ D K+M+++ PYFRP

Sbjct: 541 LTDGQGNTVDFKNTIIATSNAGFSSDAVAGE---DAKLMDKLQPYFRP 586

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5923> which encodes the amino acid sequence <SEQ ID 5924>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 551/697 (79%), Positives = 616/697 (88%), Gaps = 3/697 (0%)

Query: 5 NFYNRDPFGNMDDIFNSLMGMNGGYNSSENKRYLINGREVTPEEFSQYRQTGKLPQQLNN 64
+F +DPF NMDDIFN LM NMGGY SEN RYL+NGRE+TPEEF YRQTG+LP

Sbjct: 3 HFSGKDPFVNMDIFNQLMANMGGYRSENPRYLNGREITPEEFQHYRQTGQLPVATTKA 62

Query: 65 QNTPTNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAELARRTKNNPVLVGD 124
N+ ADSVLT+LGTNLTD +ARQ LDPVIGRNKEIQ+TAEILARRTKNNPVLVGD

Sbjct: 63 TNSQMLTPKADSVLTQLGTNLTDQARQHLLDPVIGRNKEIQETAELARRTKNNPVLVGD 122

Query: 125 AGVGKTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIKEVK 184
AGVGKTAVIEGLAQAI+NGDVPAAIKNKEI+SIDISSLEAGTQYRGSFEE IQN+I+EVK

Sbjct: 123 AGVGKTAVIEGLAQAI+VNGDVPAAIKNKEI+SIDISSLEAGTQYRGSFEETIQNLIQEVK 182

Query: 185 ETGNIILFFDEIHQILGAGSTGGDSGSKGLADILKPALSRGELTVIGATTQDEYRNTILK 244
E GNIILFFDEIHQI+GAG+T DSGSKGLADILKPALSRGELT+IGATTQDEYRNTILK

Sbjct: 183 EAGNIILFFDEIHQIVGAGATSSDSGSKGLADILKPALSRGELTLIGATTQDEYRNTILK 242

Query: 245 NAALARRFNEVKVNAPSAQDTFNILMGIRNLYEQHHNVLPDSVLKAAVDLSIQYIPQRS 304
NAALARRFNEVKVNAPSA+DTF+ILMGIRNLYEQHH++ LPD+VLKAAVD SIQYIPQRS

Sbjct: 243 NAALARRFNEVKVNAPSAEDTFHILMGIRNLYEQHHHITLPDNLKAAVDYSIQYIPQRS 302

Query: 305 LPDKAIDLDMTAAHLAAQHPVTDLKSLEKETAEQRDKQEKAVNTEDFEEALKVKTRIEE 364
LPDKAIDL+DMTAAHLAAQHPVTDLKSLEKETAEQRDKQEKAV EDKE+AL KTRIE

Sbjct: 303 LPDKAIDLDMTAAHLAAQHPVTDLKSLEKETAEQRDKQEKAVAKEDFEKALAAKTRIEE 362

Query: 365 LQNQIDNHTGQKVTATINDIAMSIERLTGVPVSNMGASDIERLKEGRLKGVIGQND 424
LQ QI+ H + Q VTAT+NDIA S+ERLTG+PVSNMG +D+ERLK + +RLK VIGQ++

Sbjct: 363 LQKQIEQHNQSNVTATVNDIAESVERLTGIPVSNMGATNDLERLKGISRLKSHVIGQDE 422

Query: 425 AVEAVARAIRRNAGFDDGNRPISFSLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMS 484
AV AVARAIRRNAGFDDG RPIGSFSLFVGPTGVGKTELAKQLA D+FGSKDAI+RLDMS

Sbjct: 423 AVAAVARAIRRNAGFDDGKRPIGSFSLFVGPTGVGKTELAKQLALDLFGSKDAIIRLDMS 482

Query: 485 EYNDRTAVSKLIGTAGYVGYDDNSNTLTERIRRNYPYSIVLLDEIEKADPQVITLLQLV 544
EYNDRTAVSKLIG TAGYVGYDDN+NTLTER+RRNPY+IVLLDEIEKADPQ+ITLLQLV

Sbjct: 483 EYNDRTAVSKLIGTTAGYVGYDDNNNTLTERVRRNPYAIIVLLDEIEKADPQIITLLQLV 542

Query: 545 DDGRLTDGQGNTINFKNTVIIATSNAGFGNEAFTGDSKDLKIMERISPYFRPEFLNRFN 604
DDGRLTDGQGNTINFKNTVIIATSNAGFG + + IM+RI+PYFRPEFLNRFN

Sbjct: 543 DDGRLTDGQGNTINFKNTVIIATSNAGFGQ---DTETSESINMDRIAPYFRPEFLNRFN 599

Query: 605 GVIEFSLSKDDLSEIVDLMLDEVNQTIGKKGIDLVDENVKSHLIELGYDEAMGVRPLR 664
+I+F+HL K+ L EIVDLML EVNQT KKG I L + ++ K+HLI+LGY+ AMG RPLR

Sbjct: 600 SIIKFNHLQKESLEEIVDLMLAEVNQTITAKKGISLTITDDAKAHLIDLGYNHAMGARPLR 659

Query: 665 RVIEQEIRDRITDYLDHTDVKHLKANLQDGQIVISE 701

R+IEQEIRDRITDYLDH +VK L+A L++GQ+VI +

-2155-

Sbjct: 660 RIIEQEIRDRTDYLDHPEVKKLQAILKEGQLVIRQ 696

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1911

A DNA sequence (GBSx2020) was identified in *S.agalactiae* <SEQ ID 5925> which encodes the amino acid sequence <SEQ ID 5926>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
10    INTEGRAL    Likelihood = -4.78    Transmembrane    8 - 24 ( 7 - 25)

----- Final Results -----
                bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9945> which encodes amino acid sequence <SEQ ID 9946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

20    >GP:AAC73364 GB:AE000134 putative enzyme [Escherichia coli K12]
        Identities = 142/307 (46%), Positives = 195/307 (63%), Gaps = 6/307 (1%)

Query: 39  KELLESKKTLLIHGALGTELESRGCDVSGKLWSAKYLIEDPAAIQTIHEDYIRAGADIVT 98
        + LL+ + L+L GA+ TELE+RGC+++ LWSAK L+E+P I+ +H DY RAGA
25    Sbjct: 8  RALLDKQDILLLDGAMATELEARGCNLADSLWSAKVLVENPELIREVHLDYYRAGAQCAI 67

Query: 99  TSTYQATLQGLAQVGVSESQTEDLIRLTVQLAKAAREQVWKS LTKEEKSERIYPLISGDV 158
        T++YQAT G A G+ E+Q++ LI +V+LA+ ARE L + ++ + L++G V
30    Sbjct: 68  TASYQATPAGFAARGLDEAQS KALIGKSVELARKAREAY---LAENPQAGTL--LVAGSV 122

Query: 159 GPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVDILALETIPNAQEAELIEL 218
        GPY A+LADGSEY G Y + + FHR R+E LLD G D+LA ET+PN E EAL EL
35    Sbjct: 123 GPY GAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAE 182

Query: 219 LAEDFPQVEAYMSFTSQDGKTSIDGSAVADLAKAIDVSPQVVALGINCSSPSLVADFLQA 278
        L +P+ A+ SFT +D + +SDG+ + D+ + PQVVALGINC + LQ
40    Sbjct: 183 LTA-YPRARAWFSFTLRDSEHLS DGTPLRDVVALLAGYPQVVALGINCIALENTTAALQH 241

Query: 279 IAEQTNKPLVTYPNSGEVYDGASQSWQSSPDHSHLTLENTSDWQKLGAQVVGCCRTTPA 338
        + T PLV YPNSGE YD S++W +H L + WQ GA+++GGCCRT PA
45    Sbjct: 242 LHGLTVLPLVVYPNSGEHYDAVSKTWHHHGEHCAQLADYLPQWQAAGARLIGGCCRTTPA 301

Query: 339 DIADLSA 345
        DIA L A
50    Sbjct: 302 DIAALKA 308

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 A related GBS gene <SEQ ID 8933> and protein <SEQ ID 8934> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 5
McG: Discrim Score:      5.48
GvH: Signal Score (-7.5): -2.64
55    Possible site: 20

```

-2156-

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program      count: 1 value: -4.78 threshold: 0.0
  INTEGRAL        Likelihood = -4.78      Transmembrane      8 - 24 ( 7 - 25)
  PERIPHERAL      Likelihood = 2.49       259
modified ALOM score: 1.46
```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF01312(412 - 1338 of 1644)
OMNI1[NT01EC0303(55 - 357 of 358) conserved hypothetical protein
%Match = 23.8
%Identity = 46.6 %Similarity = 64.3
Matches = 142 Mismatches = 107 Conservative Sub.s = 54

288 318 348 378 408 438 468 498
 LISQSFCs*FRL*GLGIAHNVLGFTSPVPHLLFSAIFITNYVTRNGDLGRFKELLESKTLILHGA LGTELESRGCDVS
 :: ||: : |:| ||: |||: |||::
 AWWPVLGWHISIQRRELRCGAGYRLRLRCAMVLISLLNPETONRSONMSONNPLRALLDKQDILLDDGAMATELEARGCNLA
 20 30 40 50 60 70 80

528 558 588 618 648 678 708 738
GKLSAKYLIEDPAAIQTIHEDYIRAGADIVTTSTYQATLQGLAQVGSSESOTEDLIRLTVLQAKAAREQVWKSILTKEEK
| | | | | : : | : : | | | | | : : | | | : : | : : | : : | : : | : : | : : |
DSLWSAKVLNPELIREVHLDDYYRGAQAICAITASYQATPAGFAARGDLDEAQS KALIGSKSVELARKARE-----AYLAEN

100 110 120 130 140 150

768 798 828 858 888 918 948 978
SERIYPLISGDVGPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLDEGVDILALETIPNAQEAELIELLAEDFFQV
: {::| ||||| |:||||| } : :: ||| |:| ||| |:| ||| |:| ||| |:| ||| |:| ||| :|:
PQAGTLLVAGSGVGPYGAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAEELLT-AYPRA
170 180 190 200 210 220 230

| 1008 | 1038 | 1068 | 1098 | 1128 | 1158 | 1188 | 1218 |
|----------------|---------------|--------------|-------------|-------------|--------------|----------|-------------|
| EAYMSFTSQDGKTI | SDGSAVADLAKA | IDVSPQVVALG | INCS | SPSLVADFLO | IAIEBOTNKPLV | TYPNSGE | VIDGASQSWQS |
| : : | : : | : : | : | : | : | | : : |
| RAWFSFTLRDSEHL | SDGTPLRDVVALL | AGYPQVVALGIN | CIALENTTAAL | QHLGLTVLPLV | VVPNSG | EHYDAVSK | TWHH |
| 250 | 260 | 270 | 280 | 290 | 300 | 310 | |

```

1248      1278      1308      1338      1368      1398      1428      1458
SPDHSHTLENTSDWQKLGAVVGGCCRTPADIALSAHLK*VKYLEEG*GKPDFLFQSTRKPAWILPNGFCFYLSMT
.:| |: |||::||| | | | | |
HGEHCAQLADYLPQWQAAGARLIGGCCRTTPADIAALKARS
          330          340          350

```

SEQ ID 8934 (GBS381) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 6; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 4; MW 66.9kDa).

Example 1912

A DNA sequence (GBSx2021) was identified in *S.agalactiae* <SEQ ID 5927> which encodes the amino acid sequence <SEQ ID 5928>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.2996(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

-2157-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1913

A DNA sequence (GBSx2022) was identified in *S.agalactiae* <SEQ ID 5929> which encodes the amino acid sequence <SEQ ID 5930>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 59
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood =-11.62    Transmembrane  176 - 192 ( 168 - 198)
        INTEGRAL    Likelihood =-11.57    Transmembrane   89 - 105 (  80 - 111)
        INTEGRAL    Likelihood =-10.03    Transmembrane  337 - 353 ( 332 - 359)
15  INTEGRAL    Likelihood = -9.87    Transmembrane  292 - 308 ( 285 - 316)
        INTEGRAL    Likelihood = -4.51    Transmembrane   58 -  74 (  52 -  78)
        INTEGRAL    Likelihood = -3.88    Transmembrane  267 - 283 ( 267 - 286)
        INTEGRAL    Likelihood = -3.08    Transmembrane  125 - 141 ( 125 - 142)
        INTEGRAL    Likelihood = -2.13    Transmembrane  212 - 228 ( 212 - 228)
20
    ----- Final Results -----
        bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
25

```

A related GBS nucleic acid sequence <SEQ ID 9377> which encodes amino acid sequence <SEQ ID 9378> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

30 >GP:CAB12034 GB:Z99105 similar to histidine permease [Bacillus subtilis]
    Identities = 221/384 (57%), Positives = 291/384 (75%), Gaps = 2/384 (0%)

    Query: 2  PVTGSFHTYATKFI SPGTGFTVAVLYWICWTVALGTEFLGAAMLQWRWFNPVPAWAFASF 61
              PVTG+FH TYA K+I PGTGFTVAVLYW+ WTVALG+EF A +LMQRWFP+ W +++
    Sbjct: 76  PVTGAFHTYAAKYIGPGTGFTVAVLYWLTWTVALGSEFTAAGLLMQRWFPHTSVMMWSAV 135

35  Query: 62  FALVIFGLNALSVRFFAEAESEFFSSIKVIAIIIFIIILGLGAMFGLVSEFEGQHKAILFTHL 121
              FAL IF LNA SV+FFAE+E +FSSIKV+AI++FI+LG AMFG++ +G A ++
    Sbjct: 136 FALFIFLLNAPSVKFFAESEFFSSIKVLAIVLFIILLGGSAMFGIPIKGEAAPMLSNF 195

40  Query: 122 TANGA-FPNGIVAVSVMLAVNYAFSGTELGIAAGETDNPKEAVPRAIKTTIGRLVVFF 180
              TA G FPNG V ++ ML+VN+AFSGTELGIAAGE+ +P + +P+AIKTT+ RL +FF
    Sbjct: 196 TAEGGLFPNGFVPILMTLSVNFAPSGTELGIAAGESVDPDKTIPKAIKTTVWRLSLFF 255

45  Query: 181 VLTIVVLASLLPMKEAGVSTAPFVDVDFDKMGIPFTADIMNFVILTALISAAGNSGLYASSR 240
              V TI VL+ L+P+++AGV +PFV VFD++G+P+ ADIMNFVILTALISA NSGLYASSR
    Sbjct: 256 VGTIFVLSGLIPIQDAGVIKSPFVAVFDRVGVPIYAADIMNFVILTALISAANSGLYASSR 315

    Query: 241 MLWSLANEGMLSKSVVKINKHGVPMRALLLSMAGAVLSLFSSIIYAADTVYLALVSIAGFA 300
              MLWSL+ E L + K+ G P AL+ SM G +LSL SS++A DTVY+ LVSI+GFA
50  Sbjct: 316 MLWSLSKEKTLHPTFAKLTSGKTPFNALVFSMIGGILSLLSSVFAPDVTYVVLVVISGFA 375

    Query: 301 VVVVWLAIPVAQINFRKEFLKE-NQLEDLSYKTPFTFVLPYITIIILLISIVGIAWDSSQ 359
              VVVVW+ I +Q FRK +++ N++ DL Y+TP P +P +L L S+VGIA+D +Q
    Sbjct: 376 VVVVWVGIAASQFMFRKRYIBAGNKVTDLKYRTPLYPFVPIAFLCLASVVGIAFDPNQ 435

55  Query: 360 RAGLYFGVPFIIFCYIYHKLRYKK 383
              R LY GVFF+ CY + ++ +K
    Sbjct: 436 RIALYCGVPFMAICYAIYYVKNRK 459

```

There is also homology to SEQ ID 4070.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1914

A DNA sequence (GBSx2023) was identified in *S.agalactiae* <SEQ ID 5931> which encodes the amino acid sequence <SEQ ID 5932>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2378(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 5642.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1915

A DNA sequence (GBSx2024) was identified in *S.agalactiae* <SEQ ID 5933> which encodes the amino acid sequence <SEQ ID 5934>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4935(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1916

A DNA sequence (GBSx2025) was identified in *S.agalactiae* <SEQ ID 5935> which encodes the amino acid sequence <SEQ ID 5936>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0530(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2159-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1917

A DNA sequence (GBSx2026) was identified in *S.agalactiae* <SEQ ID 5937> which encodes the amino acid sequence <SEQ ID 5938>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.0175(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAF63739 GB:AF236863 hypothetical GTP-binding protein
 [Lactococcus lactis]
 Identities = 142/193 (73%), Positives = 165/193 (84%)

20 Query: 6 LNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPGKTQ 65
 +NT+N +I +SAA+K YP++D PE+ALAGRSNVGKSSFINTLL RKN ARTS +PGKTQ
 Sbjct: 3 INTNNLTITISAASKKQYPENDWPEIALAGRSNVGKSSFINTLLNRKNFARTSGQPGKTQ 62

 Query: 66 LLNFYNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTNRDLRVVSLVDFRHDPS 125
 LLNFYNIDD+L FVDVPGYGYA+VSK ER KWGKMIEEYL TR+NL+ VVSLVD RH+PS
25 Sbjct: 63 LLNFYNIDDKLRFVDVPGYGYARVSKKEREKWGKMIEEYLTTRENLKAVVSLVDIRHEPS 122

 Query: 126 ADDIQMYEFLKYYEIPVIVATKADKIPRGKWNKHESIIKKLNFDKKDHFIVFSSVDRT 185
 DD+ MYEFLKYY IPVI+VATKADK+PRGKWNKHES IKK + FD D FI+FSS D+T
 Sbjct: 123 EDDLMMYEFLKYYHIPVILVATKADKVPRGKWNKHESIIKKAMKFDSTDDFIIFSSTDKT 182

30 Query: 186 GLDESWDITLSEL 198
 G++E+W IL L
 Sbjct: 183 GIEEAWTAILKYL 195

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5939> which encodes the amino acid sequence <SEQ ID 5940>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.0123(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 167/196 (85%), Positives = 183/196 (93%)

50 Query: 3 EEFLNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPG 62
 E+ LNTHNASILLSAANKSHYPQDDLPE+ALAGRSNVGKSSFINT+LGRKNLARTSSKPG
 Sbjct: 4 EQVLNTHNASILLSAANKSHYPQDDLPEIALAGRSNVGKSSFINTILGRKNLARTSSKPG 63

 Query: 63 KTQLLNFNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTNRDLRVVSLVDFRH 122
 KTQLLNF+NIDDKLRFVDVPGYGYAKVSK+ERAKWGKMIEEYL +RDNLR VVSLVD RH
 Sbjct: 64 KTQLLNFFNIDDKLRFVDVPGYGYAKVSKSERAKWGKMIEEYLTSRDNLRAVVSVDLRH 123

55 Query: 123 DPSADDIQMYEFLKYYEIPVIVATKADKIPRGKWNKHESIIKKLNFDKKDHFIVFSSV 182
 PS +DIQMY+FLKYY+IPVI+VATKADKIPRGKWNKHES +KK LNFDK D FIVFSSV
 Sbjct: 124 APSKEDIQMYDFLKYYDIPVIVATKADKIPRGKWNKHESVVKALNFDKSDTFIVFSSV 183

-2160-

Query: 183 DRTGLDESWDITLSEL 198
 +R G+D+SWD IL ++
 Sbjct: 184 ERIGIDDSWDAILEQV 199

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1918

A DNA sequence (GBSx2027) was identified in *S.galactiae* <SEQ ID 5941> which encodes the amino acid sequence <SEQ ID 5942>. This protein is predicted to be protease ClpX (clpX). Analysis of this
 10 protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9947> which encodes amino acid sequence <SEQ ID 9948>
 20 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF63738 GB:AF236863 protease ClpX [Lactococcus lactis]
 Identities = 305/395 (77%), Positives = 357/395 (90%), Gaps = 1/395 (0%)
 25 Query: 18 NVYCSFCGKSQDEVKKIIAGNGVFICNECVALSQEIIEELAEVLADLAEPKPKELLE 77
 N+ CSFCGKSQD+VKK+IAG+ V+ICNEC+ LS I++EEL EE +++ EV PKE+ +
 Sbjct: 8 NIQCSFCGKSQDDVKKMIAGSDVYICNECIELSTRILEELKEEQDSEMLEVKTPEMF 67
 30 Query: 78 ILNQYVVGQDRAKRALAVAVYNHYKRVSYTESS-DDDVDLQKSNILMIGPTGSGKTFLAQ 136
 LN+YV+GQ++AKRALAVAVYNHYKR+++T S +D++LQKSNIL+IGPTGSGKTFLAQ
 Sbjct: 68 HLNEYVIGQEKAKRALAVAVYNHYKRINFASKIAEDIELQKSNILLIGPTGSGKTFLAQ 127
 Query: 137 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGIIYVDEIDKIA 196
 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKL+QA+D+N+ERAERGIIY+DEIDKIA
 35 Sbjct: 128 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLLQASDFNTERAERGIIYIDEIDKIA 187
 Query: 197 KKGENVISITRDVSGEGVQOALLKIIIEGTVASVPPQGGRKHPNQEMIQINTKNILFIVGGA 256
 KK ENVISITRDVSGEGVQOALLKIIIEGTVASVPPQGGRKHPNQEMIQI+TKNILFIVGGA
 Sbjct: 188 KKSENVISITRDVSGEGVQOALLKIIIEGTVASVPPQGGRKHPNQEMIQIDTKNILFIVGGA 247
 40 Query: 257 FDGIEDLVKQRLGEKVIQFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFIGRLPVVAAL 316
 FDGIE++VKQRLGEK+IGFG ++K+ D SYMQEII+EDIQKFGLIPEFIGRLP+VAAL
 Sbjct: 248 FDGIEEIVKQRLGEKIIGFGANNKLSDEDSYMQEIIAEDIQKFGLIPEFIGRLPIVAAL 307
 45 Query: 317 ELLTAEDLVRILTEPRNALVKQYQTLISYDGVLEFQDALLAIADKAIERKTGARGLRS 376
 E LT EDL++ILTEP+NAL+KQY+ LL +D VELEF AL+AIA KAIERKTGARGLRS
 Sbjct: 308 ERLTEEDLIQILTEPKNALIKQYQQLLLFDNVELEFQDGMALIAKKAIERKTGARGLRS 367
 Query: 377 IIEETMLDIMFEIPSQEDVTKVRIITKAAVEGTDKP 411
 IIEE M+DIMFE+PS E++TKV IT+A V+G +P
 50 Sbjct: 368 IIEEVMMDIMFEVPSHEETTKVITTEAVVDGKAEP 402

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5943> which encodes the amino acid sequence <SEQ ID 5944>. Analysis of this protein sequence reveals the following:

55 Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2161-

bacterial cytoplasm --- Certainty=0.2711(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 378/409 (92%), Positives = 393/409 (95%), Gaps = 1/409 (0%)

Query: 9 MAGNRNDNMNVYCSFCGKSQDEVKKI IAGNGVFICNECVALSQEI I KEELAEVLADLAE 68
 MAG+R ND+ VYCSFCGKSQD+VKKI IAGN VFICNECVALSQEI I KEELAEVLADL E
 10 Sbjct: 1 MAGSRTNDIKVYCSFCGKSQDDVKKI IAGNNVFICNECVALSQEI I KEELAEVLADLTE 60

Query: 69 VPKPKELLEILNQYVVGQDRAKRALAVAVYNHYKRVS YTES- SDDVDLQKSNILMIGPT 127
 VPKPKELL++LNQYVVGQDRAKRAL+VAVYNHYKRVS+TES DDDVDLQKSNILMIGPT
 15 Sbjct: 61 VPKPKELLDVLNQYVVGQDRAKRALSVAVYNHYKRVSFTESRDDDDVDLQKSNILMIGPT 120

Query: 128 GSGKTFLAQTAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 187
 GSGKTFLAQTAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII
 20 Sbjct: 121 GSGKTFLAQTAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 180

Query: 188 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKII EGTVASVPPQGGRKHPNQEMI QINTK 247
 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKII EGTVASVPPQGGRKHPNQEMI QI+TK
 25 Sbjct: 181 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKII EGTVASVPPQGGRKHPNQEMI QIDTK 240

Query: 248 NILFIVGGAFDGI EDLVKQRLGEKVIGFGQTSRKIDDNASYMQEII SEDIQKFGLIPEFI 307
 NILFIVGGAFDGI E++VKQRLGEKVIGFGQ SRKIDDNASYMQEII SEDIQKFGLIPEFI
 30 Sbjct: 241 NILFIVGGAFDGI EEIVKQRLGEKVIGFGQNSRKIDDNASYMQEII SEDIQKFGLIPEFI 300

Query: 308 GRLPVVALELLTAEDLVRI LTEPRNALVKQYQ TLLSYDGVLEFDQDALLAIADKAIER 367
 GRLPVVALE L DL++ILTEPRNALVKQYQ LLSYDGVLEL FD++AL AIA+KAIER
 35 Sbjct: 301 GRLPVVALEQLNTSDLIQILTEPRNALVKQYQALLSYDGVLELAFDKEALEAIAANKAIER 360

Query: 368 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEGTDKPVLETA 416
 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEG KPVLETA
 Sbjct: 361 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEGKSKPVLETA 409

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1919

40 A DNA sequence (GBSx2028) was identified in *S.agalactiae* <SEQ ID 5945> which encodes the amino acid sequence <SEQ ID 5946>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1920

55 A DNA sequence (GBSx2029) was identified in *S.agalactiae* <SEQ ID 5947> which encodes the amino acid sequence <SEQ ID 5948>. Analysis of this protein sequence reveals the following:

-2162-

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4029(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9949> which encodes amino acid sequence <SEQ ID 9950>
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC33872 GB:AF055727 dihydrofolate reductase [Streptococcus pneumoniae]
 Identities = 83/162 (51%), Positives = 118/162 (72%), Gaps = 1/162 (0%)

15 Query: 25 MTKQIIAIWAEDDHGIVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLP 84
 MTK+I+AIWA+DE+ LIG LPW LP EL HFKETT+ A+LMGR TFDGM RR+LP
 Sbjct: 1 MTKKIVAIWAQDEEGLIGKENRLPWHLPAELQHFKETTLNHAILMGRVTFDGMGRRLLPK 60

20 Query: 85 RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIKTK 144
 RET+ILT++ + + DGV V+ V+ W+Q+ K L+I+GG I++AF PY + +I T
 Sbjct: 61 RETLILTRNPEEKIDGVATFQDVQSVLDWYQDQEKNYIIGGKQIFQAFEPYLDDEVIVTH 120

25 Query: 145 VHGKFKGDTYFP-DVNLSEFKVISRDYFEKDEQNAHAFTVTY 185
 +H + +GDTYFP +++LS F+ +S ++ KDE+N + FT+ Y
 Sbjct: 121 IHARVEGDTYFPEELDSLFTVSSKFKYAKDEKNPYDFTIQY 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5949> which encodes the amino acid
 sequence <SEQ ID 5950>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1214(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/160 (51%), Positives = 119/160 (74%).

40 Query: 25 MTKQIIAIWAEDDHGIVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLP 84
 MTK+IIAIWAEDD LIG+ G LPW LPKEL HFK+TT+ QA+LMGR TF+GMN + LP
 Sbjct: 1 MTKEIIAIWAEDDAGLIGIAGKLPWYLPKELEHFKTTLHQAILMGRVTFEGMNCRLPQ 60

45 Query: 85 RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIKTK 144
 R+T+++T++ +Q D V + S+E+V++W+ +KTL+I+GG + +AF Y + IIKT
 Sbjct: 61 RQTLVMTRNRDQVDEVLTMTSIEKVLWYHAQDKTLYIIGGNKVLEAFNGYFDRIIKTV 120

50 Query: 145 VHGKFKGDTYFPDVNLSEFKVISRDYFEKDEQNAHAFTVT 184
 +H +FKGDTY P+++ S F S+ ++ +D +N + FTVT
 Sbjct: 121 IHHRFKGDTYRPNLDFSHFTQESQTFYARDAKNPYDFTVT 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1921

55 A DNA sequence (GBSx2030) was identified in *S.agalactiae* <SEQ ID 5951> which encodes the amino
 acid sequence <SEQ ID 5952>. Analysis of this protein sequence reveals the following:

Possible site: 45

-2163-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1577(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAA25221 GB:M33770 thymidylate synthase (EC 2.1.1.45)
 [Lactococcus lactis]
 Identities = 215/280 (76%), Positives = 245/280 (86%), Gaps = 2/280 (0%)

Query: 1 MTKADLLFKDNITKIMSEGVSFSEQARPRYKNGEMANSKYITGAFAYDLSKGEFPITTLR 60
 MT AD +FK NI I+ GVFSE ARP+YK+G+MANSKY+TG+F YDL KGEFPITTLR
 15 Sbjct: 1 MTYADQVFKQNIQNILDNGVFSENARPKYKDGQMANSKYVTGSFVYDQLKGEFPITTLR 60

Query: 61 PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
 PIPIKSAIKE+ WIYQDQT++L+VL +KYGV YW +W +G GTIGQRYGA VKK+NII
 20 Sbjct: 61 PIPIKSAIKELMWIYQDQTSLSVLEEKYGVKYWGEWGIGD-GTIGQRYGATVKKYNIIG 119

Query: 121 KLLKQLEDNPWNRRNVISLWDYEAFEETEGLLPCAFQTMFDVRRV-NGELYLDATLTQRS 179
 KLL+ L NPWNRRN+I+LW YE FEETEGLLPCAFQTMFDVRR +G++YLDATL QRS
 30 Sbjct: 120 KLEGLAKNPWNRRNIINLWQYEDFEETEGLLPCAFQTMFDVRRREKDGQIYLDATLIQRS 179

Query: 180 NDMLVAHHINAMQYVALQMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQSEC 239
 NDMLVAHHINAMQYVALQMIAKH F W+VGKFFYF+NNLHIYDNQFEQA EL+KR SE
 25 Sbjct: 180 NDMLVAHHINAMQYVALQMIAKHFSWKVGKFFYFVNNLHIYDNQFEQANELMKRTASEK 239

Query: 240 NPKLVNLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
 P+LVLNVPDGT+FFDIKP+DF LVDY+P+KPQL+FDLAI
 30 Sbjct: 240 EPRLVLNVPDGTNFFDIKPDEFELVDYEPVKPQLKFDLAI 279

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5953> which encodes the amino acid sequence <SEQ ID 5954>. Analysis of this protein sequence reveals the following:

35 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3131(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/279 (81%), Positives = 251/279 (89%)

45 Query: 1 MTKADLLFKDNITKIMSEGVSFSEQARPRYKNGEMANSKYITGAFAYDLSKGEFPITTLR 60
 MTKAD +FK NI KI++EG SEQARP+YK+G A+SKYITGAFAYDL+KGEFPITTLR
 Sbjct: 9 MTKADQIFKANIQKIINEGSLSEQARPKYKDGRTAHSKYITGAFAYDLAKGEFPITTLR 68

50 Query: 61 PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
 PIPIKSAIKE+FWIYQDQ+N L VL KY V YWN+WEV T TIGQRYGAVVKKH+IIS
 Sbjct: 69 PIPIKSAIKELFWIYQDQSNLSLDVLEAKYNVHYWNEWEVDQTRTIGQRYGAVVKKHDIIS 128

55 Query: 121 KLLKQLEDNPWNRRNVISLWDYEAFEETEGLLPCAFQTMFDVRRVNGELYLDATLTQRSN 180
 K+LKLQ +NPWNRRNVISLWDYEAFEET+GLLPCAFQ MFDVRRV +LYLDA+LTQRSN
 Sbjct: 129 KILKQLAENPWNRRNVISLWDYEAFEETKGLLPCAFQIMFDVRRVGEDLYLDASLTQRSN 188

60 Query: 181 DMLVAHHINAMQYVALQMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQSECN 240
 D+LVAHHINAMQYVALQMIAKHFGW++GKFFYF+NNLHIYDNQF+QAQELLKRQ
 Sbjct: 189 DILVAHHINAMQYVALQMIAKHFGWKIGKFFYFVNNLHIYDNQFDQAQELLKRQPVASQ 248

Query: 241 PKLVNLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
 PKLVNLNVPD T+FFDIKPDDF L +YDP+KPQL FDLAI
 Sbjct: 249 PKLVNLNVPDRNTNFFDIKPDDFELQNYDPVKPQLHFDLAI 287

-2164-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1922

- 5 A DNA sequence (GBSx2031) was identified in *S.agalactiae* <SEQ ID 5955> which encodes the amino acid sequence <SEQ ID 5956>. This protein is predicted to be HMG-CoA synthase. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0816(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5957> which encodes the amino acid sequence <SEQ ID 5958>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1670(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 260/385 (67%), Positives = 325/385 (83%)

Query: 36 MKIGIDKIGFATSQYVLEMTDLAIARQVDPEKFSKGLLLDLSITPVTEDIVTLAASAA 95
M IGIDKIGFATSQYVL++ DLA+ARQVDP KFS+GLL++S S+ P+TEDI+TLAASAA+
Sbjct: 14 MTIGIDKIGFATSQYVLKLEDLALARQVDPKFSQGLLIESFSVAPITEDIITLAASAD 73

Query: 96 DILSDEDKETIDMVIVATESSIDQSKAASVYVHQLLEIQPFARSFEMKEACYSATAALDY 155
IL+DED+ IDIVI+ATESS DQSKA+++YVH L+ IQPFARSFE+K+ACYSATAALDY
Sbjct: 74 QILTEDRAKIDMVILATESSTDQSKASAIYVHHLVGIQPFARSFEVKQACYSATAALDY 133

Query: 156 AKLHVEKHPDSKVLVIASDIKYGKSTGESTQGAGSIAMLISQNPISILELKEDHLAQTR 215
AKLHV PDS+VLVIASDIA+YG+ S GESTQG+GSIA+L++ NP IL L ED++AQTR
Sbjct: 134 AKLHVASKPDSRVLVIASDIARYGVGSPGESTQSGSIALLVITANPRILALNEDNVAQTR 193

Query: 216 DIMDFWRPNYSDFPYVNGMFSTKQYLDMLKTTWKVYQKRFNTSLSDYAAFCFHIPFPKLA 275
DIMDFWRPNYS PYV+G++STKQYL+ L+TTW+ YQKR N LSD AA CFHIPFPKLA
Sbjct: 194 DIMDFWRPNYSFTPYVDGIYSTKQYLNCLTTWQAYQKRENQLSDLAAVCFHIPFPKLA 253

Query: 276 LKGFNKILDNNLDEQKKAELQENFEHSITYSKKIGNCYTGSlyLGLLSLENSQNLKAGD 335
LKG N I+DN + + + +L E F+ SI+YSK+IGN YTGSLYLGLLSLLENS+ L++GD
Sbjct: 254 LKGLNNIMDNTVPPEHREKLIEAFQASISYSKQIGNIYTGSlyLGLLSLENSKVLQSGD 313

Query: 336 QIAFFSYSGGAVAEIFTGQLVDGYQNKQLQSDRMDQLNKRQKITVTEYEKLFEEKTILDEN 395
+I FFSYSGGAV+E ++GQLV GY L ++R L++R +++V++YE LF+E+ LD+N
Sbjct: 314 KIGFFSYSGGAVSEFYSGQLVAGYDKMLMTNRQALLDQRTSLSVSKYEDLFYEQVQLDDN 373

Query: 396 GNANFNTYRTGTFSLSICEHQRIY 420
GNANF+ Y TG F+L +I EH+RIY

Sbjct: 374 GNANFDIYLTGKFALTAIKEHRRYIY 398

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2165-

Example 1923

A DNA sequence (GBSx2032) was identified in *S.agalactiae* <SEQ ID 5959> which encodes the amino acid sequence <SEQ ID 5960>. This protein is predicted to be HMG-CoA reductase (mvaA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 50
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.49    Transmembrane    348 - 364 ( 348 - 364)
      INTEGRAL    Likelihood = -1.33    Transmembrane    53 - 69 ( 53 - 69)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
    >GP:AAG02454 GB:AF290098 HMG-CoA reductase [Streptococcus pneumoniae]
      Identities = 266/421 (63%), Positives = 343/421 (81%), Gaps = 3/421 (0%)

20  Query: 3 KISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPFS 62
      KISW GFSKKS +ER+ L+ Q L+ + + +S+++A+Q++ENV+G .LP+S
      Sbjct: 2 KISWNGFSKKSQERLELLKAQALLSPERQASLEKDEQMSVTVADQLSENVVGTFSLPYS 61

      Query: 63 LVPDVLVNGKVQVPYVTEEPSVVAASFAAKIIRSGGFLTTHNRKMIGQVALYDVQD 122
      LVP+VLVNG+ Y VPYVTEEPSVVAAS+A+KIIKR+GGF VH R+MIGQVALY V +
25  Sbjct: 62 LVPEVLVNGQGYTVPYVTEEPSVVAASYSKIIKRAGGFTAQVHQRMIGQVALYQVAN 121

      Query: 123 SQHTKESILNQKQQLLEIANAAHPSIVKRGGGACDLTIEI---KEDFLIVYLMVDTKEAM 179
      + +E I ++K +LLE+AN A+PSIVKRGGGA DL +E + DFL+VY+ VDT+EAM
30  Sbjct: 122 PKLAQEKIASKKAELELLELANQAYPSIVKRGGGARDLHVEQIKGEPDFLVVYIHVDTQEAM 181

      Query: 180 GANMVMNTMMEALSSPLEDISKGSILMSILSNYATESLVTATCRVDLRFSLRQKEEAIKLA 239
      GANM+NTM+EAL LE++S+G+SLM ILSNYAT+SLVTA+CR+ R+LSRQK++ ++A
      Sbjct: 182 GANMLNTMLEALKPVLEELSQQGSLMILSNYATDSLVTASCRIFRYLSRQKDGREIA 241

35  Query: 240 QKMTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLS 299
      +K+ +ASQ AQ DPYRA+THNKGIFNGIDAI++ATGNDWRAIEAGAH +A +DG+Y+GLS
      Sbjct: 242 EKIALASQFAQADPYRAATHNKGIFNGIDAIIATGNDWRAIEAGAHAFASRDGRYQGLS 301

      Query: 300 RWSYKVDNDCLEGTTLTPMPVATKGGSIGINPSVHLAHDLLGRPNAKELASIIISIGLAQ 359
      W+ ++ L G +TLMPVATKGGSIG+NP V L+HDL LG P+A+ELA II+SIGLAQ
40  Sbjct: 302 CWTLDLEREELVGEMTLTPMPVATKGGSIGLNPRVALSHDLLGNPSARELAQIIISIGLAQ 361

      Query: 360 NFAALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKL 420
      NFAALKALVSTGIQ GHMKLQAKSLALLAGA E +++ +V++L+ K NLETAQ+ + L
45  Sbjct: 362 NFAALKALVSTGIQQGHMKLQAKSLALLAGASESEVAPLVERLISDKTFNLETAQRYLENL 422

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5961> which encodes the amino acid sequence <SEQ ID 5962>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 31
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3929(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 257/422 (60%), Positives = 330/422 (77%)

```

60  Query: 2 TKISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPF 61
      T ++W+GFSKK+ EER+ +E+ L +L+ + LL+ ANQM ENV+GR+ALPF

```

-2166-

Sbjct: 4 TNLNWSGFSKKTFEERLQLIEKFKLNAENLNQLKTDVLLPIQTANQMTENVLGRALPF 63

Query: 62 SLVPDVLVNGKVYQVPYVTEEPSVVAASFAAKIIKRSGGFLTTVHNRKMIGQVALYDVQ 121
 S+ PD LVNG YQ+P+VTEEPSVVAASFAAK+IKRSGGF NR+MIGQ+ LYD+

5 Sbjct: 64 SIAPDFLVNGSTYQMPFVTEEPSVVAASFAAKLIKRSGGFKAQTLNRQMIGQIVLYDID 123

Query: 122 DSQHTKESILNQKQQLLEIANAAHPISIVKRGGGACDLTIEIKEDFLIVYLMVDTKEAMGA 181
 + K +IL++ ++L+ +AN A+PSIVKRGGGA + +E K +FLI YL VDT+EAMGA

10 Sbjct: 124 QIDNAKAAILHKTFKKLIALANKAYPSIVKRGGGARTIHLEEKGEFLIFYLTVDTEAMGA 183

Query: 182 NMVNTMMEALSSPLEDISKSKSLMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLAQK 241
 NMVNTMMEAL L +SKG IM+ILSNYATESLVT +C + +R L K ++++LAQK

Sbjct: 184 NMVNTMMEALVPDLTRLKSGHCLMAILSNYATESLVTTSCEIPVRLDHDKTKSLQLAQK 243

15 Query: 242 MTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLSRW 301
 + +AS+LAQVDPYRA+THNKGIFNGIDA+V+ATGNDWRAIEAGAH YA ++G Y+GLS+W

Sbjct: 244 IELASRLAQVDPYRATTHNKGIFNGIDAVVIATGNDWRAIEAGAHAYASRNGSYQGLSQW 303

20 Query: 302 SYKVDNCLLEGTLTLPMPVATKGSIGINPSVHLAHDLLGRPNAKELASIILSIGLAQNF 361
 + D L G +TLPMP+A+KGSIG+NP+V +AHDLL +P+AK LA +I S+GLAQNF

Sbjct: 304 HFDQDKQVLLGQMTLPMPIASKGSIGLNPTVSIADLLNQPDAKTLAQLIASVGLAQNF 363

Query: 362 AALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKLT 421
 AALKAL S+GIQAGHMKL AKSLALLAGA +++I+ +V LL K +NLE A +++L

25 Sbjct: 364 AALKALTSSGIQAGHMKLHAKSLALLAGATQDEIAPLVNALLADKPINLEKAHFYLSQLR 423

Query: 422 KS 423
 +S

30 Sbjct: 424 QS 425

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1924

35 A DNA sequence (GBSx2033) was identified in *S.galactiae* <SEQ ID 5963> which encodes the amino acid sequence <SEQ ID 5964>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2355(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5965> which encodes the amino acid sequence <SEQ ID 5966>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2687(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 76/138 (55%), Positives = 100/138 (72%), Gaps = 2/138 (1%)

Query: 7 PKWEELPELDLYLDQVLLYVNQLINPKTITNDKLLTASMINNYVKHNYISKPIKKKYNRR 66
 P W++LP+LDLYLDQVLLYVNQ + ++++K LTASMINNYVKH Y++KPIKKKY ++

-2167-

Sbjct: 7 PYWKDLPLDLYLDQVLLYVNQCTDFSEVSDNKSILTASMINNYVKHGYVTKPIKKYQKQ 66

Query: 67 QVARLIVITAPKQVFAIQEISQTLLELLTADNHSEAYNGFAACMNKEE--VHDLPPVVIS 124
Q+ARLI I+ PK VF IQ+IS+ LE L A SE YN F C N++ D+PP+V

5 Sbjct: 67 QLARLIAISLFKTVFPIQDISRVLEELQAQADSESLYNTFVTCWNQKAPIEEDIPPVQV 126

Query: 125 ACQTLNLYQETQKLVLEL 142
ACQT+ Y +T L+ E+

10 Sbjct: 127 ACQTVKDYHKTIYLLQEV 144

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1925

A DNA sequence (GBSx2034) was identified in *S.agalactiae* <SEQ ID 5967> which encodes the amino acid sequence <SEQ ID 5968>. This protein is predicted to be hemolysin iii. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -9.08 | Transmembrane | 142 - 158 (140 - 165) |
| INTEGRAL | Likelihood = -6.79 | Transmembrane | 26 - 42 (19 - 44) |
| INTEGRAL | Likelihood = -5.63 | Transmembrane | 200 - 216 (196 - 217) |
| INTEGRAL | Likelihood = -5.41 | Transmembrane | 104 - 120 (102 - 121) |
| INTEGRAL | Likelihood = -3.98 | Transmembrane | 51 - 67 (49 - 69) |
| INTEGRAL | Likelihood = -1.86 | Transmembrane | 172 - 188 (169 - 188) |

----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.4630(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

A related GBS nucleic acid sequence <SEQ ID 9951> which encodes amino acid sequence <SEQ ID 9952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]

Identities = 79/204 (38%), Positives = 132/204 (63%), Gaps = 4/204 (1%)

Query: 17 EELANSITHAVGALLMLILLPITAVYSHNHFGLOAALGTSIFVTSFLMFLSSSIYHSMT 76
EE+AN+ITH +GA+L + L I +++ H A + +++ S+FL++L S++ HS+

40 Sbjct: 14 EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFSTLLHSIH 73

Query: 77 YNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYLIIFLQWGITLFGILYKIFAP 136
+ ++K + ++DHS IY+ IAG+YTP L + G LG+ ++ + W + + GI++KIF

45 Sbjct: 74 HPKVEK-LFTILDHSAIYLLIAGTYTPFLILITLRGPLEWTLAIITWLAIGGIIFKIFFV 132

Query: 137 KINDKFSVLVLYLIMGWLIVIF-IPPAITKTGPAFWGLLLAGGICYTIGALFYA-RKRPYD 194
+ K S + Y+IMGWL+I I P TG F LLLAGGI Y++GA+F+ K P++

50 Sbjct: 133 RRFKASTLCYIIMGWLIIVAIKPLYENLTGHGF-SLLLAGGILYSVGAIFFLWEKLPFN 191

Query: 195 HMIWHLFILLASILQYIGIVYFML 218
H IWHLF+L S + + +++++L

55 Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5969> which encodes the amino acid sequence <SEQ ID 5970>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

| | | | |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.51 | Transmembrane | 144 - 160 (138 - 163) |
| INTEGRAL | Likelihood = -9.87 | Transmembrane | 49 - 65 (45 - 71) |

-2168-

INTEGRAL Likelihood = -7.11 Transmembrane 198 - 214 (193 - 215)
 INTEGRAL Likelihood = -6.16 Transmembrane 102 - 118 (100 - 120)
 INTEGRAL Likelihood = -2.97 Transmembrane 20 - 36 (20 - 41)
 INTEGRAL Likelihood = -1.01 Transmembrane 167 - 183 (167 - 185)

5

----- Final Results -----

bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10

The protein has homology with the following sequences in the databases:

>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]
 Identities = 82/204 (40%), Positives = 128/204 (62%), Gaps = 4/204 (1%)

15

Query: 15 EEVANSVTHAIGAFAMLILLPISASYAYQTYDLKAAIGISIFVISLFLMFLSSTIYHSMA 74
 EE+AN++TH IGA + L I +A + A + +++ +S+FL++L ST+ HS+
 Sbjct: 14 EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFSTLLHSIH 73

20

Query: 75 YGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGITLFGILYKIFAK 134
 + V K + I+DHS IY+ IAG+YTP L + G LG+ ++ + W + + GI++KIF
 Sbjct: 74 HPKVEK-LFTILDHSAIYLLIAGTYTPFLLITLRGPLGWTLIAIWTLAIGGIIFKIFFV 132

25

Query: 135 RINEKFSMLLYIVMGWL-VVFILPVIQKTSIAFGLLMLFGGLSYTIGAVFYA-KKRPYF 192
 R K S + YI+MGWL +V I P+ T F LL L GG+ Y++GA+F+ +K P+
 Sbjct: 133 RRFIKASTLCYIIMGWLIIVAIKPLYENLTGHGFSLL-LAGGILYSVGAIFFLWEKLPFN 191

30

Query: 193 HMIWHLFILLASALQFIATFFML 216
 H IWHLF+L SA+ F + F++L
 Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/213 (71%), Positives = 181/213 (84%)

35

Query: 6 SIKLSPQLSFGEELANSITHAVGALLMLILLPTTAVYSHNHFGLQALGTSIFVTSFLFM 65
 + K S LSF EE+ANS+THA+GA MLILLPI+A Y++ + L+AA+G SIFV SLFLM
 Sbjct: 4 TFKQSLPLSFSEEVANSVTHAIGAFAMLILLPISASYAYQTYDLKAAIGISIFVISLFLM 63

40

Query: 66 FLSSSIYHSMTYNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYLEIIFLQWGIT 125
 FLSS+IYHSM Y S+ KY+LR+IDHSMIYIAIAGSYTPVALSL+ GWLGY+II LQWGIT
 Sbjct: 64 FLSSSTIYHSMAYGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGIT 123

45

Query: 126 LFGILYKIFAPKINDKFSVLVLYLIMGWLVIFFPAITTKTGPAPWGLLLAGGICYTIGAL 185
 LFGILYKIFA +IN+KFSL+LY++MGWL+FI P II KT AF L+L GG+ YTIGA+
 Sbjct: 124 LFGILYKIFAKRINEKFSMLLYIVMGWLVVFILEPVIIQKTSIAFGLLMLFGGLSYTIGAV 183

Query: 186 FYARKRPYDHMIWHLFILLASILQYIGIVYFML 218
 FYA+KRPY HMIWHLFILLAS LQ+I I +FML
 Sbjct: 184 FYAKRPYFHHMIWHLFILLASALQFIATFFML 216

50

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1926

A DNA sequence (GBSx2035) was identified in *S.galactiae* <SEQ ID 5971> which encodes the amino acid sequence <SEQ ID 5972>. Analysis of this protein sequence reveals the following:

55

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3641(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

60

-2169-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
Identities = 81/302 (26%), Positives = 157/302 (51%), Gaps = 10/302 (3%)

Query: 1 MKSAYIFFNPKSGKDEQALAKEVKSYLEHDFQDDY-VRIITPSSVEEAVAKKASEDH 59
MK A I +NP SG++ + K+ + +++ Q Y + +A AK+A+

Sbjct: 1 MKRARIINYNTSGRE---IFKKHLAQVLQKFEQAGYETSTHATTCAGDATHAAKEAALRE 57

Query: 60 IDLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNL-ALENLLN 118
DL+I GGDGTIN++ G+ PT+G++P GT N+F++AL IP+E L A + ++N

Sbjct: 58 FDLIIAAGDGTINEVVNGLAPLDNRPTLGVIPVGTINDFARALGIPREDILKAADTVIN 117

Query: 119 GHVKSVDICKVNDYMISSLTGLLADIAANVTSEMKRKLGPF AFLGDAYRILKRNRSYS 178
G + +DI +VN Y I+ G L ++ +V S++K LG A+ +L R

Sbjct: 118 GVARPIDIGQVNGQYFINIAGGGRLTELTYDVPSKLTMLGQLAYYLKGMEMLP SLRPTE 177

Query: 179 ITLAYDNNVRSRLRTRLLITMTNSIAGMPAFSPEATIDDGLFRVYTM EHIHFFKLLHLR 238
+ + YD + L L+T+TNS+ G +P++++DG+F + ++ + + +

Sbjct: 178 VEIEYDGKLFQGEIMFLVTLTNSVGGFEKLAPDSSLDGMFDMILKKNLAEFIRVAT 237

Query: 239 QFRKGFDSQAKEIKHFHTNNLTISTFKRKKSALPKVRIDGPDGDLFPVKVEVIPKALKFI 298
+G+ + I + N + ++ ++ ++ +DG+ G LP + + + + +

Sbjct: 238 MALRGEHINDQHIIYTKANRVKVNSEKM-----QLNLDGEYGGMLPGFEVNLRYRHHVV 292

Query: 299 IP 300
+P

Sbjct: 293 MP 294

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5119> which encodes the amino acid sequence <SEQ ID 5120>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4258(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/300 (57%), Positives = 229/300 (76%)

Query: 1 MKSAYIFFNPKSGKDEQALAKEVKSYLEHDFQDDYVRIITPSSVEEAVAKKASEDH 60
MK+ IF+NP SGK E LA++VK Y +H F +D V++ITP ++A LAK+A++D I

Sbjct: 1 MKTVRIFYNPNSGKKESQLARQVKDYFCQHGFSEDSVKVITPKDADQAFQLAKQAADKI 60

Query: 61 DLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNLALENLLNGH 120
DLVIPLGGDGT+NKI GG+Y GGA+ IGLVP+GTVNNF+KA++IP + AL+ +L G

Sbjct: 61 DLVIPLGGDGTLNKIIGGIYEGGAHCLIGLVPSGTVNNFAKAMHPIPLQITEALDTILTGG 120

Query: 121 VKSDICKVNDYMISSLTGLLADIAANVTSEMKRKLGPF AFLGDAYRILKRNRSYSIT 180
+K VDICK N YMISSLTGLLADIAA+VT+E KR+ GP AFL D+ RILKRNRSY+I+

Sbjct: 121 IKQVDICKANQQYMISSLTGLLADIAADVTAEEKRRFGPLAFLKDSIRILKRNRSYAIS 180

Query: 181 LAYDNNVRSRLRTRLLITMTNSIAGMPAFSPEATIDDGLFRVYTM EHIHFFKLLHLRQF 240
L N+ L+T+ LLITMTN+IAG P+FSP A DDG F+VYTM+ + FFK L H+ F

Sbjct: 181 LLSHNHRIHLKTKFLLLITMTNTIAGFSPFSGAQADDGYFQVYTMKKVSFFKFLWHINDF 240

Query: 241 RKGDFSQAKEIKHFHTNNLTISTFKRKKSALPKVRIDGPDGDLFPVKVEVIPKALKFIIP 300
++GDFS+A+EI HF N L++ +K++ +P+ RIDGD D LP+++++IPKA+ I+P

Sbjct: 241 KQGDFSQAEIISHFQANTLSLLPQAKKQAILPRTRIDGDKSDYLPQLDIIPKAVSIIVP 300

-2170-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1927

A DNA sequence (GBSx2036) was identified in *S.agalactiae* <SEQ ID 5973> which encodes the amino acid sequence <SEQ ID 5974>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:BAB10885 GB:AB010693 gene_id:K21C13.21-pir|T04769~strong
 similarity to unknown protein [Arabidopsis thaliana]
 Identities = 85/291 (29%), Positives = 150/291 (51%), Gaps = 28/291 (9%)

20 Query: 10 DQEWEPVVESEGRYHMIIVGEFCPYAQRQPQIARQLLGLDKHISISFVDDV----- 57
 D + + P ESEGRYH+ + CP+A R ++ GLD+ I+ S V +
 Sbjct: 29 DPDSQFPAESGRYHLYISYACPWACRCLSYLKI KGLDEAITFSSVHAIWGRTKETDDHRG 88

25 Query: 58 ----PSDIGLIFSQPEQVTGAKSLRDIYHLTDPTYQGPYTIPIILIDKTDNRIVCKESADL 113
 SD L ++P+ + GAKS+R++Y + P Y+G YT+P+L DK +V ES+++
 Sbjct: 89 WVFPDSDTELPGAEPDYLNAGAKSVRELYEIASPNYEGKYTVPVLDKDKLKT VVNNESS EI 148

30 Query: 114 LRLFTTDFSDLHQEDAPVLFQSQTASLIDNDIKDINKNFQSLMYKLAFLDKQADYDTYSK 173
 +R+F T+F+ + + L+ +I+ + + +YK F KQ Y+
 Sbjct: 149 IRMFNTEFNGIAKTPSLDLYPSHLRDLVINETNGWVFNGINNGVYKCGFARKQEPYNEAVN 208

35 Query: 174 EFFTFLDQKEHLLGQRPFLGDNLSEVDIHFFTPLVWRWDIAGRDL LLLNQKALEDYPNIF 233
 + + +D+ E +LG++ ++ G+ +E DI F L+R+D N++ L +YPNIF
 Sbjct: 209 QLYEAVDRCEEVLGKQRYICGNTFTADIRLFTLIRFDEVYAVHFCKNKRLLREYPNIF 268

35 Query: 234 SWAKTLYNDFNLKTLINPQSIKNNYY----LGKFGRAVRHHTIVPTGPNM 279
 ++ K +Y + + N + IK +YY + FG I+P GPN+
 Sbjct: 269 NYIKDIYQIHGMSSTVNMEHIKQHYGSHPTINPFG-----IIPHGPNI 312

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1928

A DNA sequence (GBSx2037) was identified in *S.agalactiae* <SEQ ID 5975> which encodes the amino acid sequence <SEQ ID 5976>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2647(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAB07793 GB:AB037666 hypothetical protein [Streptomyces sp.
 CL190]

-2171-

Identities = 127/331 (38%), Positives = 194/331 (58%), Gaps = 9/331 (2%)

Query: 4 RKDDHIKALKYQSHY---NSFDDIELIHSSLPKYNVNDIDLSTHFAGQSFEFFPYINAM 60
 Sbjct: 6 RKDDHVRLAIEQHNAHSGRNQFDDVSFVHHALAGIDRPDVSLSFAGISWQVPIYINAM 65

Query: 61 TGGSEK GKAVNHKLAQVAQATGIVMTGSSAALKNDE--DDSYPTTDLYPDLKLATNIG 118
 TGGSEK +N LA A+ TG+ + +GS +A +K+ D D P+ + NI
 Sbjct: 66 TGGSEKTGLINRDLATAARETGVPISGSMNAYIKDPSCADTFRVLRLDENPNNGFVIANIN 125

Query: 119 LDKPVPAESTVKAMNPIFLQVHVNMQELLMPEGEREHMRSHLKEYVDNIQCPLILK 178
 V A+ + + LQ+H+N QE MPEG+R F W +++ + P+I+K
 Sbjct: 126 ATTTVDNAQRAIDLIEANALQIHINTAQETPMPEGDRSFASWVPQIEKIAAVIDPVIVK 185

Query: 179 EVGFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGR--DRSYLNTWGQTTAQSLI 236
 EVG G+ Q+I D+G+ D+SGRGGT FA IEN R D ++L+ WQ+TA L+
 Sbjct: 186 EVGNGLSRQTILLLLADLGVQAADVSGRGGTDFARIENGRRELGDYAFHGWGQSTAAACL 245

Query: 237 NAQSMMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWK 296
 +AQ + + +LASGG+RHPLD+V+ L LGA+AVG S L + VD +I L +W
 Sbjct: 246 DAQDI--SLPVLASGGVRHPLDVVRALALGARAVGSSAGFLRTLMDGVDALITKLTTWL 303

Query: 297 EDLRMIMCALNCKKITDLRQVNYILYQLKE 327
 + L + L + DL + + +L+G+L++
 Sbjct: 304 DQLAALQTM LGARTPADLTRCDVLLHGEIRD 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5977> which encodes the amino acid sequence <SEQ ID 5978>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2823(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/329 (74%), Positives = 284/329 (86%)

Query: 1 MTNRKDDHIKALKYQSHYNSFDDIELIHSSLPKYNVNDIDLSTHFAGQSFEFFPYINAM 60
 MTNRKDDHIKALKYQS YN+FDDIELIH SLP Y+++DIDLSTHFAGQ F+FPFYINAM
 Sbjct: 31 MTNRKDDHIKALKYQSPYNAFDDIELIHSSLPYDLSIDLSTHFAGQDFDFPFYINAM 90

Query: 61 TGGSEK GKAVNHKLAQVAQATGIVMTGSSAALKNDEDDSYPTTDLYPDLKLATNIGLD 120
 TGGG+KGKAVN KLA+VA ATGIVMTGSSAALKN DDSY ++ +LKLATNIGLD
 Sbjct: 91 TGGGQK GKAVNEKLAKVAAATGIVMTGSSAALKNPNDSDSYRLHEVADNLKLATNIGLD 150

Query: 121 KPVPAESTVKAMNPIFLQVHVNMQELLMPEGEREHMRSHLKEYVDNIQCPLILKEV 180
 KPV + TV+ M P+FLQVHVNMQELLMPEGER FH W+ HL EY I P+ILKEV
 Sbjct: 151 KPVALGQQT VQEMQPLFLQVHVNMQELLMPEGERVFHTWKKHLAEYASQIPVPVILKEV 210

Query: 181 GFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGRDRSYLNTWGQTTAQSLINAQS 240
 GFGMD+ SIK A+D+GI T DISGRGGTSFAYIENQRG DRSYLN WGQTT Q L+NAQ
 Sbjct: 211 GFGMDVNSIKLAHDLGIQTTFDISGRGGTSFAYIENQRGDRSYLNDWGQTTVQCLLNAQG 270

Query: 241 MMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWKEDLR 300
 +MD+++ILASGG+RHPLDM+KC VLGA+AVGLSRTVLELVE+YP + VIAI+N WKE+L+
 Sbjct: 271 LMDQVEILASGGVRHPLDMIKCFVLGARAVGLSRTVLELVEKYPTERVIAIVNGWKEELK 330

Query: 301 MIMCALNCKKITDLRQVNYILYQLKEAN 329
 +IMCAL+CK I +L+ V+Y+LYG+L++ N
 Sbjct: 331 IIMCALDCKTIKELKGVYDLYLLYGRLLQVNV 359

-2172-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1929

A DNA sequence (GBSx2038) was identified in *S.agalactiae* <SEQ ID 5979> which encodes the amino acid sequence <SEQ ID 5980>. This protein is predicted to be phosphomevalonate kinase. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0785(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG02457 GB:AF290099 phosphomevalonate kinase [Streptococcus pneumoniae]
Identities = 170/330 (51%), Positives = 233/330 (70%), Gaps = 1/330 (0%)

Query: 1 MVKVQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLSYDMFNYTASLQPD 60
M+ V+T GKLY AGEYAIL PGQ+A++K++PIYM A F+D+Y +YSDMF++ L+P+

Sbjct: 1 MIAVKTGKLYWAGEYAILLEPGQLALIKDIPYIMRAEIAFSDSYRIYSDFDFAVDLRPN 60

Query: 61 KQYSLIQETILLMEEWLINFNGKNIKPIHLEITGKLERVGLKFGIGSSGSSVVVLTIKAMAA 120
YSLIQETI LM ++L G+N++P L+I GK+ER G KFG+GSSGSSVVVL +KA+ A

Sbjct: 61 PDYSLIQETIALMGDFLAVRGQNLRPFSLKICGKMEREGKKFGLGSSGSSVVVLVVKALLA 120

Query: 121 LYEIEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLE 180
LY + + +LLFKL++ VLLKRGDNGSMGD+ACI E L+ Y +FDR+ + +E + L

Sbjct: 121 LYNLSVDQNLFLKLTSAVLLKRGDNGSMGDLACIYAEDLVLYQSFDRQKAAAWLEENLA 180

Query: 181 QVLEAEWGYRITKIQALLEMDFLVGWMTQPSISKEMINIVKSTITQRFDDTKYQVVQLL 240
VLE +WG+ I++++ LE DFLVGWT + ++S M+ +K I Q FL +K VV L+

Sbjct: 181 TVLERDWGFFISQVKPTLECDFLVGWTKAVVSSHVMVQIKQINQNLSSSKETVVSILV 240

Query: 241 SAFKEGDKEAIKRCLEEISILLFNLHPSIYTDKIQKLKEASKGLDIVTKSSGSGGGDCGI 300
A ++G E + +E S LL L IYT L++LKEAS+ L V KSSG+GGGDCGI

Sbjct: 241 EALEQGKAEKVIEQVEVASKLLEGLSTDYITPLRLQKEASQDLQAVAKSSGAGGGDCGI 300

Query: 301 AISFN-KNDNQTLIKRWESAGIELLSKETL 329

A+SF+ ++ TL RW GIELL +E +

Sbjct: 301 ALSFDAQSSRNTLKNRWADLGIELLYQERI 330

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5981> which encodes the amino acid sequence <SEQ ID 5982>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/325 (52%), Positives = 227/325 (69%), Gaps = 2/325 (0%)

Query: 4 VQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLSYDMFNYTASLQPDQKQY 63
VQTGGKLY+ GEYAIL PGQ A++ +P+ MTA + A + L SDMF++ A + PD Y

Sbjct: 22 VQTGGKLYLTGEYAILTPGQKALIHFIPLMMAETISPAAHIQLASDMFHSKAGMTPDASY 81

-2173-

Query: 64 SLIQETILLMEEWLINFGKNIPHIHEITGKLERGLKFGIGSSGSVVVLTIKAMAALYE 123
 +LIQ T+ ++L ++P L ITGK+ER G KFGIGSSGSV +LT+KA++A Y+.
 Sbjct: 82 ALIQATVKTTFADYLGQSIDQLEPFSLIITGKMERDGGKFGIGSSGSVTLTLTKALSAYYQ 141

Query: 124 IEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLEQVL 183
 I + +LLFKL+AY LLK+GDNGSMGDIACIAY+ L++Y++FDR VS ++T PL+++L
 Sbjct: 142 ITLTPELLFKLAAYTLLKQGDNGSMGDIACIAYQTLVAYTSFDREQVSNWLQTMPLKLL 201

Query: 184 EAEWGYRITKIQALLEMDFLVGWIMQPSISKEMINIVKSTITQRFLLDDTKYQVVQ-LLSA 242
 +WGY I IQ L DFLVGWT P+IS++MI V ++IT FL T YQ+ Q + A
 Sbjct: 202 VKDWGYHIQVIQPALPCDFLVGWTKIPAIQRMIQVQTASITPAFL-RTSYQLTQSAMVA 260

Query: 243 FKEGDKKAIKRCLEEIISLLFNHPSIYTDKLOKLKEASKGLDIVTKSSGSGGGDCGIAI 302
 +EG KE +K+ L S LL LHP+IY KL L A + D V KSSGSGGGDCGIA+
 Sbjct: 261 LQEGHKEELKKS LAGASHLLKELHPAIYHPKLVTLVAACQKQDAVAKSSGSGGGDCGIAL 320

Query: 303 SFNKNDNQTLIKRWESAGIELLSKE 327
 +FN++ TLI +W+ A I LL +E
 Sbjct: 321 AFNQDARDTLISKWQEADIALLYQE 345

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1930

A DNA sequence (GBSx2039) was identified in *S.agalactiae* <SEQ ID 5983> which encodes the amino acid sequence <SEQ ID 5984>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.75 Transmembrane 20 - 36 (18 - 36)

----- Final Results -----
 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1931

A DNA sequence (GBSx2040) was identified in *S.agalactiae* <SEQ ID 5985> which encodes the amino acid sequence <SEQ ID 5986>. This protein is predicted to be mevalonate diphosphate decarboxylase. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1557(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG02456 GB:AF290099 mevalonate diphosphate decarboxylase
 [Streptococcus pneumoniae]

-2174-

Identities = 219/312 (70%); Positives = 264/312 (84%)

Query: 1 MDGKSISVKS YANIAIIKYWGKADAEKMPATSSISLTLENMYTETRLTALGKDAKDEF 60
 MD + ++V+SYANIAIIKYWGK ++M+PATSSISLTLENMYTET L+ L + DEF
 Sbjct: 1 MDREPVTVRSYANIAIIKYWGKKKEKEMVPATSSISLTLENMYTETTLSPLPANVTADDEF 60

Query: 61 YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSSGLSALVKACND 120
 YI+G LQN+ EH KMS I+DR+R GFV+I+T NNMPTAAGLSSSSSSGLSALVKACN
 Sbjct: 61 YINGQLQNEVEHAKMSKIIDRYRPAGEGFVRIDTQNNMPTAAGLSSSSSSGLSALVKACNA 120

Query: 121 FFGTNLSQS QLAQEAKFASGSSSRSFPGPVAAWDKDSGDIYKVHTNLDLAMIIMLVLNDR 180
 +F L +S QLAQEAKFASGSSSRSF+GP+ AWDKDSG+IY V T+L LAMIIMLVL DK+
 Sbjct: 121 YFKLGLDRS QLAQEAKFASGSSSRSFYGPLGAWDKDSGEIYPVETDLKLAMIIMLVLEDKK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNNDFOKVGQLTERNALAMHSTT 240
 KPISSR+GMK+C ETSTTF++WVRQSE+DYQDML+YLK NDF K+G+LTE+NALAMH+TT
 Sbjct: 181 KPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYKENDFAKIGELTEKNALAMHATT 240

Query: 241 KTATPAFSYLTETETKAMDVVKKLREKGHECYTMDAGPNVKVLCRLQDLEALAAILEKD 300
 KTA+PAFSYLT+ +Y+AM V++LREKG CY+TMDAGPNVKV C +DLE L+ I +
 Sbjct: 241 KTASPAFSYLT DASYEAMAFVRLREKGEACYFTMDAGPNVKVFCQEKDLEHLSEIFGQR 300

Query: 301 YRIIVSTTKELA 312
 YR+IVS TK+L+
 Sbjct: 301 YRLIVSKTKDLS 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5987> which encodes the amino acid sequence <SEQ ID 5988>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 221/313 (70%), Positives = 258/313 (81%)

Query: 1 MDGKSISVKS YANIAIIKYWGKADAEKMPATSSISLTLENMYTETRLTALGKDAKDEF 60
 +D I+V SYANIAIIKYWGK + KMIP+TSSISLTLENM+T T ++ L A D+F
 Sbjct: 1 VDPNVITVTSYANIAIIKYWGKENQAKMIPSTSSISLTLENMFTTTSVSFLPDTATSDQF 60

Query: 61 YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSSGLSALVKACND 120
 YI+G+LQND EH K+SAI+D+FRQ FVK+ET NNMPTAAGLSSSSSSGLSALVKAC+
 Sbjct: 61 YINGILQNDDEHTKISAIIDQFRQPGQAFVKMETQNNMPTAAGLSSSSSSGLSALVKACDQ 120

Query: 121 FFGTNLSQS QLAQEAKFASGSSSRSFPGPVAAWDKDSGDIYKVHTNLDLAMIIMLVLNDR 180
 F T L Q LAQ+AKFASGSSSRSFPGPVAAWDKDSG IYKV T+L +AMIMLVLN +
 Sbjct: 121 LFDTLQDQKALAQKAKFASGSSSRSFPGPVAAWDKDSGAIYKVETDLKMAMIMLVLNAK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNNDFOKVGQLTERNALAMHSTT 240
 KPISSREGMK+C +TSTTF++WV QS DYQ ML YLK N+F+KVGQLTE NALAMH+TT
 Sbjct: 181 KPISSREGMKLCRDTSTTFDQWEQSAIDYQHMLTYLKTNNFEKVGQLTEANALAMHATT 240

Query: 241 KTATPAFSYLTETETKAMDVVKKLREKGHECYTMDAGPNVKVLCRLQDLEALAAILEKD 300
 KTA PFSYLT+E+Y+AM+ VK+LR++G CY+TMDAGPNVKVLCCL +DL LA L K+
 Sbjct: 241 KTANPPFSYLTKE SYQAMEAVKELRQEGFACYFTMDAGPNVKVLCLEKDLAQLAERLGKN 300

Query: 301 YRIIVSTTKELAD 313
 YRIIVS TK+L D
 Sbjct: 301 YRIIVSKTKDLPD 313

-2175-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1932

A DNA sequence (GBSx2041) was identified in *S.agalactiae* <SEQ ID 5989> which encodes the amino acid sequence <SEQ ID 5990>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1512(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5991> which encodes the amino acid sequence <SEQ ID 5992>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1117(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 182/290 (62%), Positives = 223/290 (76%)

```

Query: 1  MKEKFGIGKAHSKIILMGEHVVYGYPAIAIPLKNIEVTCLIEEAPQLIALDMTDPLSTA 60
          M E G GKAHSKIIL+GEH+VVYGYPAIA+PL +IEV C I A + + D D LSTA
Sbjct: 6  MNENIGYGKAHSKIILIGEHAVVYGYPAIALPLTDIEVVCHIFPADKPLVDFDYDTLSTA 65

Query: 61  IFAALDYLGKTSSKIAYHIESQVPERRMGSSAAVAIAAIRAVFDYFDEDLEADLLECLV 120
          I+A+LDYL + IAY I SQVP++RGMGSSAAV+IAAIRAVF Y E L DLLE LV
Sbjct: 66  IYASLDYLQRLQEPIAYEIVSQVPQKRMGSSAAVSTAAIRAVFSYQEPLESDDLLEILV 125

Query: 121 NRAEMIAHSNPSGLDAKTCLSENTIKFIRNIGFSTVPMHLNAYLVIADTGIHGHTKEAVD 180
          N+AE+IAH+NPSGLDAKTCLS++ IKFIRNIGF T+ + LN YL+IADTGIHGHT+EAV+
Sbjct: 126 NKAETIIAHTNPSGLDAKTCLSDHAIKFIRNIGFETIEIALNGYLI+IADTGIHGHTREAVN 185

Query: 181 KVKSSGEAVLPFLKELGYLAASEDAIHKSQSLGSLMTKAHQSLKQLGVSSLEADHLV 240
          KV E LP+L +LG L +A E AI++ + +G LMT+AH +LK +GVS +AD LV
Sbjct: 186 KVAQFEETNLPYLAKLGALTQALERAINQKNKVAIGQLMTQAHSALKAGVSISKADQLV 245

Query: 241 EVAISCGALGAKMSGGGLGGCIIALVKEKREAERLSQQLEREGAVNTWTE 290
          E A+ GALGAKM+GGGLGGC+IAL K AE++S +L+ EGAVNTW +
Sbjct: 246 EAALRAGALGAKMTGGGLGGCMIALADTKDMAEKISHRLKEEGAVNTWIQ 295

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1933

A DNA sequence (GBSx2042) was identified in *S.agalactiae* <SEQ ID 5993> which encodes the amino acid sequence <SEQ ID 5994>. This protein is predicted to be a histidine protein kinase. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood =-13.43 Transmembrane 12 - 28 (4 - 33)

-2176-

INTEGRAL Likelihood = -9.29 Transmembrane 163 - 179 (157 - 191)

----- Final Results -----

5 bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAF79919 GB:AF039082 putative histidine protein kinase
[Lactococcus lactis]
Identities = 78/315 (24%), Positives = 154/315 (48%), Gaps = 33/315 (10%)

Query: 101 SDRQIKNYAKRIVSQNSHSGHITYNFSTYSYLLKKVKGNDYLVVFLDTTNQYLDNQRLLQ 160
+++QI N + + +N + + Y + T S + V++ + Q +

15 Sbjct: 84 NEKQI-NTIQTVSVKNPYGDNWHYRYLTTSQFIITNSDGTVPVYVQIFSNDQIQDAMS 142

Query: 161 LSIWM---SLVSFIVFMVIVSV-LSGRVILPFVANYEKQRRFITNAGHELKTPLAIISAN 216
++W+ ++++F + VI+S+ L+ + P +A YEKQ+ F+ NA HEL+TPLAI+

20 Sbjct: 143 RAMWVIVTTMITFWILSVIISLYLANWTLKPILAAYEKQKEFVENASHELRTPLAILQNR 202

Query: 217 NELV-----EMSGESEWTKSTNDQIQRLTGLINGMVSLAR-----FEEQPDISM---- 261
EL+ + +SE + +++ + L + +++LAR E +P +

Sbjct: 203 LELLFQKPTATIIDQSENISESLSEVRNMRLTSLNLLNLARRDSGIKIEPEPTTATYFEN 262

25 Query: 262 VDLDFSHITKDAEDFKGPIIKDGKDFIMSIQPGIHVKAEEKSLFELVTLLVDNANKYCD 321
+ + +T++A + F G + +G V ++ + +L+T+L DNA KY D

Sbjct: 263 IFNSYEMLTENAGKKFSGNLKLEGT-----VNLDQALIKQLLTILFDNALKYTD 311

Query: 322 PMGTVTVKLSRSSRLRRAKLEVSNTYKNGKIDIDYSKFFERFYREDESHNNKSGYGIGLS 381
G ++V + ++ V++ + D D K F+RF+R D++ +K G G+GLS

30 Sbjct: 312 SEGEISVDVIKNGGF--LTFADNNGEGISDEDEKKKIFDRFFRVDKARTRQKGGGLGLGLS 369

Query: 382 IVTSLVHLFKGSIDV 396
+ +V + G I V

35 Sbjct: 370 LAKQIVEAYNGKITV 384

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5751> which encodes the amino acid sequence <SEQ ID 5752>. Analysis of this protein sequence reveals the following:

40 Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -11.30 Transmembrane 18 - 34 (13 - 42)
INTEGRAL Likelihood = -10.35 Transmembrane 170 - 186 (163 - 199)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 233/410 (56%), Positives = 303/410 (73%), Gaps = 1/410 (0%)

Query: 1 MFRNRLRLRFIGIAALAILVVLFSVVGVLNSANHYQTKNEIYRVLTILADNNGRIPNKLEF 60
MF +R+RFI IA++AI ++L S+VG++N+A YQ++ EI R+L +++ N G++P E

55 Sbjct: 10 MFNRIIRIRFIMIASIAIFIILSSVIGIINTARCYQSQQEINRILHLISSNKGKLPGTTES 69

Query: 61 SKELGDDLSTDAIFQFRYFSARTDAKGNVTSFDSRNIFEVSDRQIKNYAKRIVSQNSHSG 120
SK LG LS D++ QFRY+S +A G++ S ++ NI + + + +A+ G

Sbjct: 70 SKRLGTLKLSQFRYYSVIFNANGHLLSSNTANISALDREEAQYFARLFAKSGEKEG 129

60 Query: 121 HITYNFSTYSYLLKKVKGNDYLVVFLDTTNQYLDNQRLLQLSIWMSLVSFIVFMVIVSVL 180
+ S YSYL+ ++ + LVV LDTT + LL +S+ ++ FI F+V+VS+

Sbjct: 130 SYRHQDSVYSYLITQLPNEEKLVVILDITTFYFRSVGDLAVSVMLAFGGFIFFVVLVSLF 189

Query: 181 SGRVILPFVANYEKQRRFITNAGHELKTPLAIISANNELVEMSGESEWTKSTNDQIQRL 240

-2177-

SG VI PFV NYEKQRRFITNAGHELKTPLAIISSANNELVE+M+GESEWTKST+DQ++RL
 Sbjct: 190 SGMVVKPFVQNYEKQRRFITNAGHELKTPLAIISSANNELVELMTGESEWTKSTSDQVKRL 249

Query: 241 TGLINGMVSLARFEEQPDISMVDLDFSHITKDAEDFKGPIIKDGKDFIMSIQPGIHVKA 306
 TGLIN M++LAR EEQPD+ + +DFS I +DAAEDFK ++KDGF F ++IQP I +KA
 Sbjct: 250 TGLINQMITLARLEEQPDVVLHMFSAIAQDAEDFKSLVLKDGKRFDLTIQPNIMIKA 309

Query: 301 EEKSLFELVTLTLLVDNANKYCDPMGTVTVKLSRSSRLR-RAKLEVSNTYKNGKDIDYSKFF 359
 EEKSLFELVT+LVDNANKYCDP G V V L+ R R RAKLEVSNTY GK IDYS+FF
 Sbjct: 310 EEKSLFELVTILVDNANKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLEGKSIDYSRFF 369

Query: 360 ERFYREDESHNNKSGYGIGLSIVTSLVHLFKGSIDVNYKHDTITFVIYI 409
 ERFYREDESHN+K+ GYGIGLS+ S+V LFKG+I VNYK+D I F + I
 Sbjct: 370 ERFYREDESHNSKEKGYGIGLSMAESMVKLFGKTITVNYKNDIVFTVVI 419

SEQ ID 5994 (GBS273) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 14; MW 46kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 5; MW 71kDa).

GBS273-GST was purified as shown in Figure 208, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1934

A DNA sequence (GBSx2043) was identified in *S.agalactiae* <SEQ ID 5995> which encodes the amino acid sequence <SEQ ID 5996>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2181(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1935

A DNA sequence (GBSx2044) was identified in *S.agalactiae* <SEQ ID 5997> which encodes the amino acid sequence <SEQ ID 5998>. This protein is predicted to be two-component response regulator (trcR).

Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2503(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9379> which encodes amino acid sequence <SEQ ID 9380> was also identified.

-2178-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
Identities = 71/183 (38%), Positives = 120/183 (64%), Gaps = 3/183 (1%)

5 Query: 9 RVLIAEDEEQMSRVLSTAISHQGYVVDVAYDGQTAIDLANQAYDVMVMDVMPVKTGIE 68
R+LI EDE++++RVL + H+GY D A+ G ++ +A+D++++DVM+P +G+E
Sbjct: 3 RILIIEDEKKIARVLQLELEHEGYETDAAFSGSDGLETFQAHAWDLVLLDVMLPELSGLE 62

10 Query: 69 AVKEIRQSGNKSHIIMLTAMAEIDDRVTGLDAGADDYLTKEPFLKELLARLRSMRRL- 127
++ IR + + II+LTA I D+V+GLD GA+DY+TKPF ++ELLAR+R+ R ++
Sbjct: 63 VLRRIRMTDPVTPILLTARNIPDKVSGLDGANDYITKPFIEELLARVRACLRVQT 122

15 Query: 128 -DFTPNVLSLGRVTLVSGEQELQCEN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVV 185
+ + L +T++ +++Q N TI L KE ++L FF+ N + LS +Q+ +VW
Sbjct: 123 RERVEDTLMFQELTINEKTRDVQRNETIELTPKEFELLVFFIKNGQVLSREQILTQNV 182

Query: 186 GAD 188
G D
Sbjct: 183 GFD 185

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5999> which encodes the amino acid sequence <SEQ ID 6000>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/185 (67%), Positives = 151/185 (81%)

35 Query: 8 MRVLIAEDEEQMSRVLSTAISHQGYVVDVAYDGQTAIDLANQAYDVMVMDVMPVKGTGI 67
M++L+AEDE QMS VL+TA++HQGY VDV ++GQ AID A NAYD+M++D+MMP+K+GI
Sbjct: 1 MKILLAEDEWQMSNVLTAMTHQGYVDVVFNGQEAIDKAKONAYDIMILDIMMPIKSGI 60

40 Query: 68 EAVKEIRQSGNKSHIIMLTAMAEIDDRVTGLDAGADDYLTKEPFLKELLARLRSMRRL- 127
EA+KEIR SGN SHIIMLTAMAEI+DRVTGLDAGADDYLTKEPFLKELLARLRSM RR+E
Sbjct: 61 EALKEIRASGNC SHIIMLTAMAEINDRVTGLDAGADDYLTKEPFLKELLARLRSMERRVE 120

Query: 128 DFTPNVLSLGRVTLVSGEQELQCEN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVVGA 187
FTD VI. VPI+++ RORT. N TRLA KE K++AF MTAN K T. TRL T.L.A.TKRW

-2179-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAB05604 GB:AP001513 unknown conserved protein [Bacillus halodurans]
 Identities = 67/182 (36%), Positives = 111/182 (60%), Gaps = 4/182 (2%)

Query: 17 LEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRKQLPVTVD SALKTIRDS 76
 L++ + +I + + + Y EH+ R+K+ +S++ K +R+ T++S + +RD

10 Sbjct: 29 LQELNTKIDILKQEFQYIHDYNPIEHVSSRVKSPESIVNKIQRRGNDFTLESIRENVRDI 88

Query: 77 IGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDYIQHVKNPNGYRSYHVILEIDTPYPDCL 136
 G+RI C F +DIY + E++ D +V KDYI++ KPNGYRS H+IL I P +

15 Sbjct: 89 AGIRITCSFESDIYTLSEQLMQQHDISVVETKDYIKNPKPNGYRSLHLILSI---PIFM 144

Query: 137 GNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIE NPERIVRELKRCADEMASVDLTMQT 196
 + Y+E+Q+RTIA D WASLEH++ YK++ PE +++ELK A+ A +D M+

Sbjct: 145 SDRVQDVYVEVQIRTIAMDFWASLEHKIYYKYNKNVPEHLLKELKDAESAALLDQKMEK 204

20 Query: 197 IR 198
 I+

Sbjct: 205 IQ 206

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6003> which encodes the amino acid
 sequence <SEQ ID 6004>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1057(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 127/206 (61%), Positives = 162/206 (77%)

Query: 3 TNIYGDYGRYLP LILEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRKQL 62
 ++IY + YLPL+L+ + I EN K+K ETG+KLYEH RIK+ SMIEK+RKQL

40 Sbjct: 11 SSIYSGFEVYLPLVLQTITDVIIAENIKSKKETGFKLYEHFTSR IKSEAS MIEKCQRKQL 70

Query: 63 PVTVD SALKTIRDSIGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDYIQHVKNPNGYRSY 122
 P+T SALK I+DSIG+RIICGF++DIY++++ +K+ + EKDYI + KPNGYRSY

Sbjct: 71 PLTSKSALKIKKDSIGIRIICGFIDDIYRMVDLLKSIPGMSVNT EKDYILNAKPNGYRSY 130

45 Query: 123 HVILEIDTPYPDCLGNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIE NPERIVRELKR 182
 H+ILE++T +PD LG G Y+IE+QLRTIAQDSWASLEHQMKYKH + N E I RELKR

Sbjct: 131 HLILELETHFPDILGEKKGCFIEVQLRTIAQDSWASLEHQMKYKHQVANAEMITRELKR 190

Query: 183 CADEMASVDLTMQTIRQLIESGTTKE 208
 CADE+AS D+TMQTIRQLI+ T++E

50 Sbjct: 191 CADELASCDVTMQTIRQLIQETTEE 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1937

A DNA sequence (GBSx2046) was identified in *S.agalactiae* <SEQ ID 6005> which encodes the amino acid sequence <SEQ ID 6006>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

-2180-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3250(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA37193 GB:X53013 ORF1 (AA 1 - 384) [Lactococcus lactis]
 Identities = 30/55 (54%), Positives = 37/55 (66%)

Query: 1 MEFYKTLKRKFINDADTIFIEQSQFEIFYIETDHNSSSHVLDYQSQKEFEK 55
 ME +YKTLKR+ INDA ++ EIF YIET +N+ H LDYQS K+FEK
 Sbjct: 327 MESFYKTLKRELINDAHFETRAEATQEIFKYIETYYNTKWMHSGLDYQSPKDFEK 381

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6007> which encodes the amino acid sequence <SEQ ID 6008>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3065(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 31/59 (52%), Positives = 39/59 (65%)

Query: 1 MEFYKTLKRKFINDADTIFIEQSQFEIFYIETDHNSSSHVLDYQSQKEFEKIITN 59
 ME +YKTLKR+ +NDA I+Q+Q EIF Y ET +N H L Y S EFEKI+T+
 Sbjct: 13 MEAFYKTLKRELVDNAHFATIKQAQLEIFKYSETYYNPKRLHSALGYLSPVEFEKIVTH 71

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1938

A DNA sequence (GBSx2047) was identified in *S.agalactiae* <SEQ ID 6009> which encodes the amino acid sequence <SEQ ID 6010>. This protein is predicted to be R5 protein. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.98 Transmembrane 30 - 46 (29 - 51)
 INTEGRAL Likelihood = -2.76 Transmembrane 967 - 983 (966 - 985)

----- Final Results -----

bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8935> which encodes amino acid sequence <SEQ ID 8936> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 SRCFLG: 0
 McG: Length of UR: 2
 Peak Value of UR: 2.44
 Net Charge of CR: 2
 McG: Discrim Score: 0.78
 GvH: Signal Score (-7.5): -0.0599995

-2181-

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 40
 ALOM program count: 0 value: 7.37 threshold: 0.0
 PERIPHERAL Likelihood = 7.37 194
 modified ALOM score: -1.97

*** Reasoning Step: 3

Rule gp01

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

LPXTG motif: 944-948

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8936 (GBS200) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 3; MW 107.4kDa), in Figure 169 (lane 4; MW 122kDa) and in Figure 238 (lane 11; MW 122kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 3; MW 132kDa).

Purified Thio-GBS200-His is shown in Figure 244, lane 9.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1939

A DNA sequence (GBSx2048) was identified in *S.agalactiae* <SEQ ID 6011> which encodes the amino acid sequence <SEQ ID 6012>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3919 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9953> which encodes amino acid sequence <SEQ ID 9954> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16108 GB:Z99124 similar to transcriptional regulator (MarR family) [Bacillus subtilis]
 Identities = 30/114 (26%), Positives = 59/114 (51%), Gaps = 3/114 (2%)

Query: 29 DVEHLAGPQGHLMVLYKHPDKDMSIKAVEEILHISKSVASNLVSRMEKNGFTAI VPSKT 88
 D++ G +LV +Y++P + + + E++ + ++ A+ +K++E GFI +P +
 Sbjct: 25 DLDLTRGQYLYLVR-IYENPG--IIQEKLAEMIKVDRTTAARA I KKLEMQGFIQKLPDEQ 81

Query: 89 DKRVKYLTLTHLGKKKATQFEIFLEKLHSTMLAGITKKEIRT TTKKVIRTLAKNM 142
 +K++K L+ T GKK E L+G T EE T ++ + KN+
 Sbjct: 82 NKKIKKLFPTKGGKVYPLLRREGEHSTEVLSGFTSEEKETISALLHRVRKNI 135

-2182-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6013> which encodes the amino acid sequence <SEQ ID 6014>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4175(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 27/64 (42%), Positives = 46/64 (71%)

Query: 3 MENPLQKARILVNQLEKYLDHYAKEYDVEHLAGPQGHLMVLYKHPDKMSIKAVEEILH 62

M + R L++Q+E+ D AK+YDVEHLAGPQG++++L KH ++++ +K +E+ L

Sbjct: 1 MSQVIGDLRELIHQIEQISDEIAKKYDVEHLAGPQGYVLVFLAKHQEQEIFVKDIEKQLR 60

Query: 63 ISKS 66

I +S

Sbjct: 61 IFQS 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1940

A DNA sequence (GBSx2049) was identified in *S.agalactiae* <SEQ ID 6015> which encodes the amino acid sequence <SEQ ID 6016>. This protein is predicted to be 5'-nucleotidase family protein. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -2.66 Transmembrane 668 - 684 (665 - 684)

----- Final Results -----

bacterial membrane --- Certainty=0.2062(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12747 GB:Z99108 similar to 5'-nucleotidase [Bacillus subtilis]

Identities = 178/535 (33%), Positives = 270/535 (50%), Gaps = 55/535 (10%)

Query: 28 DQGVGVQVIGVNDVDFHGALDNTGTANMPDGKVANAGTAAQLD---AYMDDAQKDFKQTNPNG 84

+ V ++++ +ND HG +D ++ DG GT ++D AY+ + + + K

Sbjct: 586 EHVPLRILSMNDLHGKIDQQVELDL-DGNGTVDTGTFGRMDYAAAYLKEKKAEEKN----- 639

Query: 85 ESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVEYGTIGNHEFDEGLAEYNRIVTGKA 144

S+ V AGDM+G S S LLQDEPTV+ + + GT+GNHEFDEG E RI+ G

Sbjct: 640 -SLIVHAGDMIGGSSPVSSLLQDEPTVELMEDIGFDVGTGNGHEFDEGTDELLRILNG-G 697

Query: 145 PAPDSNINNITKSYPEAAKQEVIVVANVIDKVNKQIPYNWKPYAIAKNIPVNNKSVNVGFI 204

P +++P +V AN ++ +P+ +N + V V FI

Sbjct: 698 DHPKGTSGYDQGNFP-----LVCANC-----KMKSTGEFPFLPAYDIINVEGVPVAFI 744

Query: 205 GIVTKDIPNLVLRKNYEQYEFIDEAETIVKYAKELQAKNVKAIIVLAHVPATSKNDIAEG 264

G+VT+ +V+ + + EF DEA + K A+EL+ K VKAI VLAH+ A + G

Sbjct: 745 GVVTSQAAGVMPEGIKNIEFTDEATAVNKAAEELKKGKVKAIIVLAHMSAEQNGNAITG 804

Query: 265 EAAEMMKVNVQLFPENSVDIVFAGHNHQYTNGLVGKTRIVQALSQKAYADVRGVLDITDT 324

E+A++ K ++ +D++FA HNHQ NG V IVQA GKA V +D T

-2183-

Sbjct: 805 ESADLANKT-----DSEIDVIFAHHNQVNVNGEVLNGKLIVQAFEYGKAIGVVDVEIDKTT 859

Query: 325 QDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIVKQVTEAKIGTAESVSMITRSVDQD 384
+D ++ SA+++ V K AI+ + TI + + +G A V + S D D

5 Sbjct: 860 KDIVK-KSAEIVYVDQSKI EPDVSASAILKKYETIAEPTISEVVGEAAVDMEGGYSNDGD 918

Query: 385 NVSPVGSLLITEAQLAIARKSWPDIDFAMTNNGGIRADLLIKPDGTITWGAQAQVQPFNGI 444
+P+G+LI + A + DFA+ N GGIR L G ITWG +QPFNG+

10 Sbjct: 919 --TPLGNLIADGMRAAMK-----TDFALMNGGGIREAL---KKGPITWGDLYNIQPFNGV 968

Query: 445 LQVVEITGRDLYKALNEQYDQKQNFLLQIAGLRVYTYTDNKEGGEETPFKVVKAYKSNGEE 504
L +EI G+DL + +N Q I+G +TYT +KE G+ K+ ++G E

Sbjct: 969 LTKLEIKGKDLREIINAQISPVFGPDYSISG--FTYTWDKETGKAVIDMKM-----ADGTE 1021

15 Query: 505 INPDAKYKLVINDFLFGGGDGFASFRNAKLLGAINP-----DTEVFMAXITDLEK 554
I PDA Y L +N+F+ A ++ LLG NP D E + Y+ ++

Sbjct: 1022 IQPDATYTLTVNNFMATATG--AKYQPIGLLGK-NPVTGPEDEATVEYVKSFDE 1073

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1607> which encodes the amino acid
20 sequence <SEQ ID 1608>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.67 Transmembrane 662 - 678 (661 - 679)

INTEGRAL Likelihood = -2.02 Transmembrane 19 - 35 (18 - 35)

----- Final Results -----

bacterial membrane --- Certainty=0.2869(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 415/688 (60%), Positives = 517/688 (74%), Gaps = 21/688 (3%)

Query: 1 MKKKIILKSSVLGLVAGTSIMFSSVFADQVGVQVIGVNDFHGALDNTGTANMPDGKVANA 60
MKK ILKSSVL ++ +++ + V ADQV VQ +GVNDFHGALDNTGTATA P GK+ NA

35 Sbjct: 14 MKKYFILKSSVLSILTSFTLLVTDVQADQVDVQFLGVNDFHGALDNTGTATYTPSGKIPNA 73

Query: 61 GTAAQLDAYMDDAQKDFKQTNPNGESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVE 120
GTAAQL AYMDA+ DFKQ N +G SIRVQAGDMVGASPANS LLQDEPTVK FN M E

40 Sbjct: 74 GTAAQLGAYMDDAEIDFKQANQDGT SIRVQAGDMVGASPANSALLQDEPTVKVFNKMKFE 133

Query: 121 YGTLGNHEFDEGLAEYNRIVTGKAPAPDSNINNITKSYPHAAKQEIVVANVIDKVNKQI 180
YGTLGNHEFDEGL E+NRI+TG+AP P+S IN+ITK Y HEA+ Q IV+ANVIDK K I

45 Sbjct: 134 YGTLGNHEFDEGLDEFNRIMTGQAPDESTINDITKQYEHEASHQITIVIANVIDKKTCDI 193

Query: 181 PYNWKPYAIKNIPVNNKSVNVGFIGIVTKDIPNLVLRKNYEQYFLDEAETIVKYAKELQ 240
PY WKPYAIK+I +N+K V +GFIG+VT +IPNLVL++NYE Y+FLD AETI KYAKELQ

Sbjct: 194 PYGWKPYAIKDIAINDKIVKIGFIGVVTETIPNLVLKQNYEHYQFLDVAETIAKYAKELQ 253

50 Query: 241 AKNVKAIIVVLAHVPATSKNDIAEGEAAEMMKVNQLFPENSVDIVFAGHNNHQYTNGLVGK 300
++V AIVVLAHVPATSK+ + + E A +M+KVNQ++PE+S+DI+FAGHNNHQYTNG +GK

Sbjct: 254 EQHVHAIIVVLAHVPATSKDGVVDHEMATVMEKVNQIYPEHSIDIIIFAGHNNHQYTNGTIGK 313

Query: 301 TRIVQALSQGKAYADVRGVLDTDTQDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIV 360
TRIVQALSQGKAYADVRG LDTDT DFI+TPSA V+AVAPG KT ++DI+AI++ AN IV

55 Sbjct: 314 TRIVQALSQGKAYADVRGTLDTDTNDFIKTPSANVVAVAPGIKTENSIDIKAIINHANDIV 373

Query: 361 KQVTEAKIGTAESVSMITRSVDQDNVSPVGSLLITEAQLAIARKSWPDIDFAMTNNGGIRA 420
K VTE KIGTA S I+++ + D SPVG+L T AQL IA+K++P +DFAMTNNGGIR+

60 Sbjct: 374 KTVTERKIGTATNSSTISKTENIDKESPVGNLATTATLTIKKTFTPTVDFAITNNGGIRS 433

Query: 421 DLIKPDGTITWGAQAQVQPFNGILQVVEITGRDLYKALNEQYDQKQNFLLQIAGLRVYTY 480
DL++K D TITWGAQAQVQPFNGILQV+++TG+ +Y LN+QYD+ Q +FLQ++GL YTY

Sbjct: 434 DLVVKNDRTITWGAQAQVQPFNGILQVIQMTGQHIYDVLNQQYDENQTYFLQMSGLTYTY 493

65

-2184-

Query: 481 TDNKEGGEETPFKVVKAYKSNNGEENPDACYKLVINDFLFGGGDGFASFRNAKLLGAINP 540
 TDN +TPFK+VK YK NGEEIN Y +V+NDFL+GGDGF++F+ AKL+GAIN
 Sbjct: 494 TDNDPKNSDTPFKIVKVYKDNNGEENLNTTITVTVVNDFLYGGDGFSAFKKAKLIGAINP 553

5 Query: 541 DTEVFMAYITDLEKAGKKVSPNNKPKIYVTMKNVNETITQNDGTHSIIKKLYLDRQNI 600
 DTE F+ YIT+LE +GK V+ K YVT + + T + G HSII K++ +R GN
 Sbjct: 554 DTEAFITTYITNLEASGKTVNATIKGVKNYVTSNLESSTKVNSAGKHSIISKVFRNRDGN 613

10 Query: 601 VAQEIVSDTLNQTKSKSTKINPVTTIHKQLHQFTAINPMRNYGKPSNSTTVKSKQLPKT 660
 V+ E++SD L T++ + + T +N T+ S LP T
 Sbjct: 614 VSSEVISDLLTSTENTNNSLGKKET-----TTNKNTISSSTLPIT 653

Query: 661 NSEYGQSFLMSVFG-VGLIGIALNTKKK 687
 Y S +M++ + L G+ KK+
 15 Sbjct: 654 GDNYKMSPIMTILALISLGLNAFIKKR 681

SEQ ID 6016 (GBS328) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 4; MW 73kDa). The GBS328-His fusion product was purified (Figure 213, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 268), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1941

A DNA sequence (GBSx2050) was identified in *S.agalactiae* <SEQ ID 6017> which encodes the amino acid sequence <SEQ ID 6018>. This protein is predicted to be peptide deformylase (def-2). Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.70 Transmembrane 55 - 71 (55 - 74)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB09662 GB:Z96934 peptide deformylase [Clostridium
 beijerinckii]
 Identities = 71/136 (52%), Positives = 96/136 (70%)

40 Query: 1 MIKPIVRDTFFLQOKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIINV 60
 MIKPIV+D FL QKS+ A++ D+ + DL +TL AN +CVG+AAANMIG KR+++ V
 Sbjct: 1 MIKPIVKOILFLGQKSEATKNDMVVIDDLIDTLRANLEHCVGLAANMIGVKKRILVFTV 60

45 Query: 61 GITNLVMFNPVVVAKSDPYETEESCLSLVGCSTQRYCHITISYRDINWKEQQIKLTDFP 120
 G + M NPV++ K PYETEESCLSL+G R T+RY I ++Y D N+ +++ F
 Sbjct: 61 GNLIVPMINPVLKKEKPYETEESCLSLIGFRKTKRYETIEVTVLDRNFNKKKQVFNGFT 120

Query: 121 AQICQHELDHLEGILI 136
 AQI QHE+DH EGI+I
 50 Sbjct: 121 AQIIQHEMDHFEGIIII 136

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6019> which encodes the amino acid sequence <SEQ ID 6020>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

-2185-

INTEGRAL Likelihood = -3.61 Transmembrane 55 - 71 (55 - 73)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/136 (56%), Positives = 103/136 (75%)

10 Query: 1 MIKPIVRDITFLLQKKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIIINV 60
MI+ I+ D F LQKK+Q+A + D+ + +DLQ+TL + C+GMAANMIG KR++I+++
Sbjct: 1 MIREIITDHFLLQKKAQVAKKEDLWIGQDLQDTLAFYRQECLGMAANMIGEQRIVIVSM 60

15 Query: 61 GITNLVMFNPVVAKSDPYETEESECLSLVGCSTQRYCHITISYRDINWKEQQIKLTDFP 120
G +LVMFNPV+V+K Y+T+ESCLSL G R TQRY IT+ Y D NW+ +++ LT
Sbjct: 61 GFIDLVMFNPVMVSKGIYQTKESCLSLSGYRKTQRYDKITVEYLDHNWRPKRLSLTGLT 120

20 Query: 121 AQICQHELDHLEGILI 136
AQICQHELDHLEGILI
Sbjct: 121 AQICQHELDHLEGILI 136

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1942

A DNA sequence (GBSx2051) was identified in *S.agalactiae* <SEQ ID 6021> which encodes the amino acid sequence <SEQ ID 6022>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2880(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05820 GB:AP001514 NADP-specific glutamate dehydrogenase
[Bacillus halodurans]

Identities = 298/444 (67%), Positives = 362/444 (81%), Gaps = 2/444 (0%)

40 Query: 7 YVASVLEKVKQNEHEEEFLQAVEEVFESLVVPFDKYPQYIEENLLERLVEPERVISFRV 66
YV V E VK++N +E EF QAV+EVF+SL+PV K+PQY+++ +LER+VEPERVISFRV
Sbjct: 16 YVQHVVYETVKRRNPNEHEFHQAVKEVFDSSLPLVVKHPQYVKQAILERIVEPERVISFRV 75

45 Query: 67 PWVDDKGQVQVNRGVRVQFSSAIGPYKGLRFHPTVTQSIVKFLGFQIFKNSLTGLPIG 126
PWVDD+G VQVNRG+RVQF+SA+GPYKGLRFHP+V SI+KFLGFQIFKN+LTG PIG
Sbjct: 76 PWVDDQGNVQVNRGFRVQFNSALGPYKGLRFHPSVNASIIKFLGFQIFKNALTGQPIG 135

50 Query: 127 GKGKGSNFPDPKGKSDNEVMRFTQSFMTLQKYIGPDLDPAGDIGVGGREIGYLYGQYKR 186
GKGKGS+FDPKGKSD E-MRF+QSF+EL YIGPD+DVPAGDIGVG +EIGY++GQYK+
Sbjct: 136 GKGKGSDFDPKGKSDGEIMRFSQSFMSSELSNYIGPDIDVPAGDIGVGAKEIGYMFQYKK 195

55 Query: 187 L-NGYQNGVLTGKGLTYGGLARTEATGYGAVYFAKEMLAARGQDLTGKVALVSGSGNVA 245
+ G++ GVLTKG+ YGGLAR EATGYG VYF +EM+ G G +VSGSGNV+
Sbjct: 196 MRGGFEAGVLTGKIGYGGSLARKEATGYGTVYFVEEMIKDHGFSFAGSTVVVSGSGNV 255

Query: 246 IYATEKLQELGATVVAVSDSSGYVYDPDGIDLETLKQIKEVERARIVKYTEKHPKANFTP 305
IYA EK +LGA VVA SDS GYVYD +GIDL+T+K++KEVER RI +Y +HP A++
Sbjct: 256 IYAMEKAMQLGAKVVACSDSGGYVYDKNGIDLTQTVKRLKEVERKRISSEYVNEHPHAHYVQ 315

60 Query: 306 ADQGSISIKADLAFPCATQNELDEEDAKLLVENGVLAVTEGANMPSTLGAIKVFQKAGV 365

-2186-

G IWS+ D+A PCATQNELDE A +L+ NGV AV EGANMPSTL A+ FQ+ GV
 Sbjct: 316 GCSG-IWSVPCDIALPCATQNELDEAAATMLIANGVKAVGEGANMPSTLQAVHTFQEHGV 374

Query: 366 AFGPAKAANAGGVAVSALEMAQNSSRRRAWTFEEVDQELQRIIMKTIFVNASEAADEFQDSG 425
 F PAKAANAGGV+VSALEMAQNS+R AWTTFEEVD +L IMK I+ + +AA+ + SG
 Sbjct: 375 LFAPAKAANAGGVSVSALEMAQNSTRLAWTTFEEVDKLYEIMKNYRESIKAELYEASG 434

Query: 426 NLVLGANIAGFLKVAQAMSAQGIV 449
 NLV+GANIAGF+KVA AM + G+V

Sbjct: 435 NLVVGANIAGFVKVADAMISHGVV 458

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 1943

A DNA sequence (GBSx2052) was identified in *S.agalactiae* <SEQ ID 6023> which encodes the amino acid sequence <SEQ ID 6024>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.55 | Transmembrane | 61 - 77 (55 - 87) |
| INTEGRAL | Likelihood = -7.70 | Transmembrane | 177 - 193 (175 - 202) |
| INTEGRAL | Likelihood = -7.06 | Transmembrane | 99 - 115 (95 - 122) |
| INTEGRAL | Likelihood = -5.89 | Transmembrane | 42 - 58 (40 - 60) |
| INTEGRAL | Likelihood = -3.08 | Transmembrane | 160 - 176 (159 - 176) |
| INTEGRAL | Likelihood = -2.44 | Transmembrane | 124 - 140 (122 - 144) |

----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.4418(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

A related GBS nucleic acid sequence <SEQ ID 9955> which encodes amino acid sequence <SEQ ID 9956> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1944

A DNA sequence (GBSx2053) was identified in *S.agalactiae* <SEQ ID 6025> which encodes the amino acid sequence <SEQ ID 6026>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

| | | | |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.72 | Transmembrane | 152 - 168 (147 - 192) |
| INTEGRAL | Likelihood = -5.47 | Transmembrane | 267 - 283 (264 - 288) |
| INTEGRAL | Likelihood = -4.30 | Transmembrane | 171 - 187 (169 - 192) |
| INTEGRAL | Likelihood = -2.13 | Transmembrane | 67 - 83 (67 - 83) |
| INTEGRAL | Likelihood = -0.32 | Transmembrane | 493 - 509 (493 - 509) |

----- Final Results -----

| | | | |
|--------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.5288(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |

-2187-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB69752 GB:AL137187 putative ABC transporter [Streptomyces coelicolor A3(2)]
Identities = 269/611 (44%), Positives = 392/611 (64%), Gaps = 31/611 (5%)

Query: 9 RLWSYLTRYKATLFLAIFLKVLSFMSILEPFILGLAITELTANLV--DMAKG----- 59
RL S +ATLF + V+S ++++ P ILG A + A +V DM G

10 Sbjct: 27 RLVSQFRPERATLFTLLACVVSVGLNVVGPILGRATDLVFAGIVGRDMPGATKEQVL 86

Query: 60 -----VSGAELNVPYIAGILIIYFRGVFVELGSYGSNYFMTTVV 99
V G ++ + +L++ L + + V

Sbjct: 87 ATMREHGDGNVADMLRSTDFVPGQIDFGAVGEVLLALATFAVAGLLMAVATRLVNRAV 146

15 Query: 100 QKSIRDIRHDLNRKINKVPVSFYFDKHQFGDMLGRFTSDVETVSNALQQSFLQIINAFLSI 159
+++ +R D+ K++++P+SYFDK Q G++L R T+D++ + LQQS Q+IN+ L+I

Sbjct: 147 NRTMFRRLREDVQTKLSRLPLSYFDKQRGEVLSRATNDIDNIGQTLQQSMGQLINSLITI 206

20 Query: 160 ILVVVMVLYLNVPLAMIIICIPVTFYSAQAILKRSQPYFKEQAKILGELNGFVQEKLTG 219
I V+ M+ Y++ LA++ + +P+++ A + KRSQF F +Q + G+LN ++E TG

Sbjct: 207 IGVLAMMFYVSWILALVALVTVPVSFVATRVGKRSQPQFVQWRSTGQLNAHIEEMYTG 266

Query: 220 FNIKLYGREEASSQEFRDITDNLRHVGFKASFISGIMMPVLNSISDFIYLIIFVGGGLQ 279
++K++GR+E S+++F + D L GFKA F SGIM P++ +S+ Y+++A VGGL+

25 Sbjct: 267 HALVKVFGROESAKQFAEQNDALYEAGFKAQFNSGIMQPLMCMVSNLNVVLVAVVGGLR 326

Query: 280 VIAGTLTIGNMQAFVQYVWQISQPVQTTITQLAGVLQSAKSSLERIFEVL D-EEEEANQVT 338
V +G L+IG++QAF+QY Q S P+ + +A ++QS +S ER+FE+LD EE+ A+ +

30 Sbjct: 327 VASGQLSIGDVQAFIQYSRQFSMPLTQVASMALVQSGVASAERVFELLDAAEQSADPIP 386

Query: 339 EKLSDLTGQVSFHFVDFHYSFDPKPLIRDFNL DVEPGQMI AIVGPTGAGKTTLINLLMRF 398
DL G+V V F Y P+KPLI D +L VEPG +AIVGPTGAGKTTL+NLMLRF

Sbjct: 387 GARPEDLRGRVLEHVSFRYDPEKPLIEDLSLKVEPGHTVAIVGPTGAGKTTLVNLLMRF 446

35 Query: 399 YDVSEGAITVDGHDIRHLSRQDFRQFGMVLQDAWLYEGTIKENLRF G-NLEASDEDIVA 457
Y+VS G IT+DG DI +SR + R GMVLQD WL+ GTI EN+ +G + E + +I

Sbjct: 447 YEVSGGRITLDGVDIAKMSRDEL RAGIGMVLQDTWLFGGTIAENIAYGASREVTRGEIEE 506

40 Query: 458 AAKAANVDHFIRTLPGGYNMVMNQESSNISLGQKQLLTIA RALLADPKILILDEATSSVD 517
AA+AA+ D F+RTLP GY+ V++ E + +S G+KQL+TIARA L+DP IL+LDEATSSVD

Sbjct: 507 AARAAHADRFRVRLPDGYDTVIDDEGTGVSAGEKQLITIA RAFLSDPVILVLDEATSSVD 566

Query: 518 TRLELLIQKAMKLMGRTSFVIAHRLSTIQEADNILVLKDGQIIEQGNHQLLADKGFY 577
TR E+LIQKAM KL GRTSFVIAHRLSTI++AD ILV++DG I+EQG H +LL G Y

45 Sbjct: 567 TRTEVLIQKAMAKLAHGRTSFVIAHRLSTIRDADTILVMEDGAIVEQGAHTELLTADGAY 626

Query: 578 YELYNSQFSNS 588
LY +QF+ +

50 Sbjct: 627 ARLYKAQFAEA 637

There is also homology to SEQ IDs 160 and 6546.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1945

55 A DNA sequence (GBSx2054) was identified in *S. agalactiae* <SEQ ID 6027> which encodes the amino acid sequence <SEQ ID 6028>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

60 INTEGRAL Likelihood = -10.88 Transmembrane 242 - 258 (235 - 263)
INTEGRAL Likelihood = -9.82 Transmembrane 159 - 175 (129 - 177)
INTEGRAL Likelihood = -9.71 Transmembrane 52 - 68 (49 - 77)

-2188-

INTEGRAL Likelihood = -8.49 Transmembrane 134 - 150 (129 - 158)
 INTEGRAL Likelihood = -1.17 Transmembrane 272 - 288 (272 - 289)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces
 coelicolor A3(2)]
 Identities = 226/565 (40%), Positives = 342/565 (60%), Gaps = 1/565 (0%)

15 Query: 6 SYLKRYPNWLWLDLLGAMLFVTVILGMPTALAGMIDNGVTGKDRGTGVYLWTFIMFIFVVL 65
 +YL+ Y + L + L L +PT A +ID GV KGD + + +M +
 Sbjct: 8 TYLRPYKKPIALLVALQFLQTCASLYLPTLNAHIIDEGVVKGDSGYILSYGALMIGISLA 67

20 Query: 66 GIIGRITMAYASSRLTTTMRDMRNDMYAKLQEQYSHHEYEQIGVSSLVTRMTSDTFVLMQ 125
 ++ I + +R + RD+R ++ +Q +S E G SL+TR T+D +
 Sbjct: 68 QVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDVQQVQM 127

25 Query: 126 FAEMSLRLGLVTPMVMIFSVVMILITSPSLANLVAVAMPLLVGVLVVAIKTKPLSERQQ 185
 A M+ L + P++ + +VM L L+ ++ +P+L + + K +PL + Q
 Sbjct: 128 LALMTFTLMVSAPIMCVGGIVMALGLDVLGVLGVVPVLAICVTLIVRKLRLFRKMQ 187

30 Query: 186 TMLDKINQYVRENLTGLRVVRAFARENFSQKFOVANQRYTDTSTGLFKLTGLTEPLFVQ 245
 LD +N+ +RE +TG RV+RAF R+ ++ Q+F+ AN T+ + G L L P+ +
 Sbjct: 188 VRLDTVNRVLRQITGNRVIRAFVRDEYEQQRFKANTELTEVALGTGNLLALMFPVVM 247

35 Query: 246 IIIAMIVAIVWFALDPLQGAIGKIDLVAFIEYSFHALFSFLLFANLFTMYPRMVSSHR 305
 ++ +A+VWF + G ++IGDL AF+ Y + S ++ +F M PR V + R
 Sbjct: 248 VVNLSSIAVWVFGAHRIDSGGMQIGDLTAFLAYLMQIVMSVMMATFMFMMPRAEVCAR 307

40 Query: 306 IREVMMPISINPNTGVTDTKLKGHLEFDNVTFAYPGETESPVLHDISFKAKPGETIAF 365
 I+EV++ S+ P VT+ + GHLE F YPG E PVL I A+PGET A
 Sbjct: 308 IQEVLETSSVPPVAPVTELRRHGHLEIREAGFRYPG-AEEPVLRHIDLVARPGETTAV 366

45 Query: 366 IGSTGSGKSSLVNLIPRFYDVTGKILVDGVDVRDYNLKSRLQKIGFIPQKALLFTGTIG 425
 IGSTGSGKS+L+ L+PR +D T G++LV+GVDVR + K+L + + +PQK LF GT+
 Sbjct: 367 IGSTGSGKSTLLGLVPRLFDATDGEVLVNGVDVRTVDPKTLAKVVSLLVPQKPYLFAGTVA 426

50 Query: 426 ENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLSGGQKQRLSIARAV 485
 NL+YG DAT ++L A+ ++QAKEF+ + + +A+GG+N+SGGQ+QRL+IAR +
 Sbjct: 427 TNLRYGNPDATDEELWHALAVAQAKEFVSELEGGDLAPIAQGGTNVSGGQKQRLAIARTL 486

55 Query: 486 VKDPDLIYFDDSFSDYKTDATLRLARLKEVTGDSVTLIVAQRVGTIMDADQIIVLDEGE 545
 V+ P++Y+YFDDSFSDY TDA LRA L + T ++TV+IVAQRV TI DAD+I+VLDEG
 Sbjct: 487 VQRPEIYLFDDSFSDYATDAALRAELAQETAETVIVIVAQRVATIRADRIIVLDEGR 546

60 Query: 546 IVGRGTHAQLIENNAIYREIAESQL 570
 +VG G H +L+ +N YREI SQL
 Sbjct: 547 VVGVRHHELMADNETYREIVLSQL 571

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4985> which encodes the amino acid
 55 sequence <SEQ ID 4986>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

60 INTEGRAL Likelihood = -16.24 Transmembrane 155 - 171 (145 - 176)
 INTEGRAL Likelihood = -7.48 Transmembrane 130 - 146 (122 - 150)
 INTEGRAL Likelihood = -5.04 Transmembrane 13 - 29 (12 - 30)
 INTEGRAL Likelihood = -5.04 Transmembrane 56 - 72 (52 - 75)
 INTEGRAL Likelihood = -4.14 Transmembrane 239 - 255 (238 - 259)
 INTEGRAL Likelihood = -1.70 Transmembrane 269 - 285 (269 - 288)

65 ----- Final Results -----

-2189-

bacterial membrane --- Certainty=0.7496(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 175/511 (34%), Positives = 296/511 (57%), Gaps = 3/511 (0%)

Query: 59 MFIFVVLGIIGRITMAYASSRLTTTMRDMRNDMYAKLQEYSHHEYEQIGVSSSLVTRMTS 118
 + I +LG++ +++++ + DMR + K+Q++S+ E +LV R+T+
 10 Sbjct: 56 LLIIALLGLMSGAINTVLAAKIAQGVSAADMREKTFRKIQDFSYANIEAFNAGNLVVRLLTN 115

Query: 119 DTFVLMQFAEMSLRLGLVTPMVMIFSVMILITSPSLAWLVAVAMPLLVGVLVYVAIKTK 178
 D + M ++ P++ I + +M + T P L W++ V + L+ ++ V +
 15 Sbjct: 116 DINQIQSLVMMMFQILFRLPILFIGAFIMAVQTFPQLWWVIVVMVILIALIMGLVMRQMG 175

Query: 179 PLSEKQQTMLDKINQYVRENLTGLRVVRAFARENFSQKQFQVANQRYTDTSTGLFKLTGL 238
 P + Q ++DKIN+ +ENL G+RVV++F +E Q KF+ + + + L
 Sbjct: 176 PRFGKFQRLMDKINRIAKENLRGVRVVKSFVQEQQYTKFKETSNDLLALNLSIGYGFSL 235

20 Query: 239 TEPLFVQIIIIAMIVAIVWFALDPLQGAIGDLVAFIEYSFHALFSFLLFANLFTMYPR 298
 +P + + + + ++ IG++ +F+ Y +FS ++ ++ R
 Sbjct: 236 MQPALMLVSYLAVVVSINVSTMVETDPTVIGNIASFMTYMMQIMFSIIVGSMGMQVSR 295

25 Query: 299 MVVSSHRIREVMDMPISINPTEGVTDTKLKGHLEFDNVTFAYPGETESPVLHDSFKAK 358
 VS RIR+++ ++ E + + G + FD+V+F YP + E P L ISF +
 Sbjct: 296 AFVSMARIRQILSTEPAMTFENE--KEETISGSIVFDDVSFTYPNDDE-PTLKHISFAIE 352

30 Query: 359 PGETIAFIGSTGSGKSSLVNLIPRFYDVTLGKILVDGVDVRDYNLKSRLQKIGFIPQKAL 418
 PG+ + +G+TGSGKS+L LIPR +D G+IL+ G ++ + +LRQ + + QKA+
 Sbjct: 353 PQQMGVIGVATGSGKSTLAQLIPRLFDPODQIILGGKPIKTLSTLRSQSVSIVLQKAI 412

35 Query: 419 LFTGTIGENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLGGQKQR 478
 LF+GTI +NL+ G A A ID +++A I+QAKEFI+ +E+ + E GSNLGGQKQR
 Sbjct: 413 LFSGTIADNLRQGSAAKADIDAMQKAAQIAQAKEFIDRMSRYESQVEERGSNLGGQKQR 472

40 Query: 479 LSIARAVVKDPDLXIIFDDSFSAIDYKTDATLRLARLKEVTGDSVTLIVAQRVGTIMDADQI 538
 LSIAR V+ P + I DDS SALD K++ ++ L +T +IVAQ++ +++ AD+I
 Sbjct: 473 LSIARGVINHPKILILDDSTSALDAKSEKRVQEAALSHKLEGTITTVIVAQKISSVVKADKI 532

40 Query: 539 IVLDEGEIVGRGTHAQLIENNAIYREIAESQ 569
 +VLD+G+++G GTHA+L+ NNAIYREI E+Q
 Sbjct: 533 LVLDDGQLIGEGTHAELVANNAIYREIYETQ 563

There is also homology to SEQ IDs 72 and 6552.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1946

A DNA sequence (GBSx2055) was identified in *S.agalactiae* <SEQ ID 6029> which encodes the amino acid sequence <SEQ ID 6030>. Analysis of this protein sequence reveals the following:

50 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA51784 GB:X73368 ORF 18.3 [Salmonella typhimurium]

-2190-

Identities = 58/162 (35%), Positives = 92/162 (55%), Gaps = 8/162 (4%).

Query: 1 MIIRPIIKNDQAVAQLIRQSLRAYDL--DKPDTAYSDPHLDHLTSYYEKIEKSGFFVIE 58
 + +R I D+ A+A++IRQ Y L DK T +DP+LD L Y + + ++V+E
 5 Sbjct: 9 LTVRRITADNAAIARVIRQVSAEYGLTADKGYTV-ADPNLDELYQVYSQ-PGAAYWVVE 66

Query: 59 ERDEIIGCGGFGPLKNL---IAEMQKVYIAERFRGKGLATDLVKMIEVEARKIGYRQLYL 115
 + ++G GG PL I E+QK+Y RG+GLA L M AR+ G+++ YL
 10 Sbjct: 67 QNGCVVGGGGVAPLSCSEPDICELQKMYFLPVIRGQGLAKKLALMALDHAREQGFKRCYL 126

Query: 116 ETASTLSRATAVYKHMGYCALSQPIANDQGHTAMDIWMIKDL 157
 ET + L A A+Y+ +G+ +S+P+ GH ++ M+KDL
 Sbjct: 127 ETATFLREAIALYERLGFHEHISEPL-GCTGHVDCEVRMLKDL 167

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1947

20 A DNA sequence (GBSx2056) was identified in *S.agalactiae* <SEQ ID 6031> which encodes the amino acid sequence <SEQ ID 6032>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12566 GB:Z99108 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 269/625 (43%), Positives = 397/625 (63%), Gaps = 11/625 (1%)

35 Query: 1 MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
 MS + L K+ GDKT+F ++SF I +RIG+IG NGTGK+TLL VI+G +
 Sbjct: 1 MSILKAENLYKTYGDKTLFDHISFHIEENERIGLIGPNGTGKSTLLKVIAGLESIE--EG 58

40 Query: 61 PFSSANDYKIAYLKQEPDFDSDTILDTVLSSDLREMAIKEYELLNHY-----EESKQ 115
 + + ++ +L Q+P+ QT+L+ + S + M ++EYE L E +Q
 Sbjct: 59 EITKSGSVQVEFLHQDPELPAGQTVLEHIYSGESAVMKTLEYEKALYELGKDPENBQRQ 118

45 Query: 116 SRLEKVMAMDSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDAD 175
 L A+MD+ +AW + KTVLSKLG+ D+ V ELSSG ++RV +A+ L+ AD
 Sbjct: 119 KHLAQAQAKMDANNAWDANTLAKTVLSKLGVDVTKPVNELSGGQKKRVAIAKNLIQPAD 178

Query: 176 LLLLDEPTNHLIDITIAWLTNFKNSKKTIVLFITHDRYFLDNVATRI FELDKAQITEYQG 235
 LL+LDEPTNHL +TI WL +L V+ +THDRYFL+ V RI+EL++ + Y+G
 50 Sbjct: 179 LLILDEPTNHLDNETIEWLEGYLSQYPGAVMLVTHDRYFLNRVTNRIYELERGSlyTYKG 238

Query: 236 NYQDYVRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQARINRFQNLKNDLH 295
 NY+ ++ RAE++ + K++ L ++ELAW+R +AR+TKQ+ARI+R + LK
 Sbjct: 239 NYEVFLEKRAEREAQAEQKETKRONLLRRELAWLRGAKARSTKQKARIDRVETLKEQTG 298

55 Query: 296 QTSDDTLEMTFTFSRIGKKVINFEVNSFSYPDKSILKDFNLLIQNKDRIGIVGDNVGVK 355
 S S L+ + R+GK+VI ENV +Y + ++ FN L+ +RIGI+G NG+GK
 Sbjct: 299 PQSSGS-LDFAIGSHRLGQVIEAENVMIAYDGRMLVDRFELVIPGERIGIIGPNGIGK 357

60 Query: 356 STLLNLIVQDLQPDGNSIGETIRVGYFSQQLHMDGSKRVINYLQEVADDEVKTSVGTI 415
 +TLLN + PD G+++IG+T+R+GY++Q M+G +VI+Y++E A+ VKT+ G

-2191-

Sbjct: 358 TTLLNALAGRHTPDGGDITIGQTIVRIGYYTQDHSEMNGELKVIDYIKETAEEVVKIADGDM 417

Query: 416 SVTE-LLEQFLFPRSTHGTQIAKLSGGEKKRLYLLKILIEKPNVLLLEPTNDLDIATLT 474
E +LE+FLFPRS T I KLSGGEK+RLYLL++L+++PNVL LDEPTNDLD TL+

5 Sbjct: 418 ITAEQMLERFLFPRSMQQTYYIRKLSGGEKKRLYLLQVLMQEPNVFLDEPTNDLDTETLS 477

Query: 475 VLENFLQGGFPGPVITVSHDRYFLDKVANKIIAFEDND-IREFFGNYYTDYLDEKAFNEQNN 533
VLE+++ F G VITVSHDRYFLD+V ++I FE N I F G+Y+DY++E +

10 Sbjct: 478 VLEDYIDQFPFPGVVITVSHDRYFLDRVVDRLIVFEGNGVISRFQGSYSYDMEESKAKKAAP 537

Query: 534 EVISKKESTKTSREKQSRKRMSYFEKQEWATIEDDIMILENTITRIENDMQTCGSDFTRL 593
+ + +E T + K+ RK++SY ++ EW IED I LE ++E D+ GSDF ++

Sbjct: 538 KP-AAEEKTAAEAPKKRKLKLSYKDQLEWDGIEDKIAQLEEKHEQLEADIAAGSDFGKI 596

15 Query: 594 SDLQKELDAKNEALLEKYDRYEYLS 618
+L E E L DR+ LS

Sbjct: 597 QELMAEQAKTAELEAAMDRTWTELS 621

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6033> which encodes the amino acid sequence <SEQ ID 6034>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2591(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 467/624 (74%), Positives = 535/624 (84%), Gaps = 3/624 (0%)

Query: 1 MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
MS FLV+ LTK+VGDKTVF ++SFIH DRIGIIGVNGTGKTTLLDV+SG LGFDGD S

35 Sbjct: 1 MSHFLVEKLTKTVGDKTVFQDISFIHDFDRIGIIGVNGTGKTTLLDVLSGRLGFDGDHS 60

Query: 61 PFSSANDYKIAYLKQEPDFDSSQTILDTVLSSDLREMAIKEYELLLNHYEESKQSRLEK 120
PFS ANDYKIAYL Q+P+F+D+ ++LDTVLS+D++ + LI++YELL+ +Y E KQ LE

Sbjct: 61 PFSKANDYKIAYLTDQPEFNDAASVLDTVLSADVKAIQIRQYELLMANYTEDKQESLES 120

40 Query: 121 VMAEMDSLDAWSIESSEVKTIVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDADLLLLL 180
+M+EMD LDAWSIES+VKTIVLSKLGITDL+ VG+LSGG+RRRVQLAQVLL ADLLLLL

Sbjct: 121 LMSEMDRLDAWSIESDVKTIVLSKLGITDLEQKVGDLSGGMRRRVQLAQVLLGAADLLLLL 180

45 Query: 181 EPTNHLDDITIAWLTNFLKNSKKTIVLFITHDRYFLDNVATRIFELDKAQITEYQGNQDY 240
EPTNHLDDITIAWLT +LK +KKTIVLFITHDRYFLD+VATRIFELDKA +TEYQGNQDY

Sbjct: 181 EPTNHLDDITIAWLTTYLKTAKKTIVLFITHDRYFLDHVATRIFELDKAGLTEYQGNQDY 240

Query: 241 VRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQARINRFQNLKNDLHQTSDT 300
VRL+AEQDERDAA+LHKKKQLYKQELAWMRTQPQARATKQARINRF +LK ++HQ S

50 Sbjct: 241 VRLKAEQDERDAANLHKKKQLYKQELAWMRTQPQARATKQARINRFSDLKKEVHQDSSA 300

Query: 301 SDLEMTFETSRIGKKVINFENVFSYPDKSILKDFNLLIQNKDRIGIVGDNGVGKSTLLN 360
LEMTFETSRIGKKVI+FE++SF+Y D+ ++KDFNL+IQNKDRIGIVGDNGVGKSTLLN

55 Sbjct: 301 DKLEMTFETSRIGKKVIHFEDLSFAYGDRQLIKDFNLLIQNKDRIGIVGDNGVGKSTLLN 360

Query: 361 LIVQDLQPDSCGNVSIGETIRVGYFSQQLHNMDGSKRVINYLQEVADDEVKTSVGTTSVTEL 420
+I DL+P SG + IG+TIRVGYFSQQL +MD +KRVINYLQEVADDEVKTSVGTTS++EL

Sbjct: 361 IINGDLKPTSGKLDIGDTIRVGYFSQQLKMDKETKRVINYLQEVADDEVKTSVGTTSISEL 420

60 Query: 421 LEQFLFPRSTHGTQIAKLSGGEKKRLYLLKILIEKPNVLLLEPTNDLDIATLT/VLENFL 480
LEQFLFPRS+HGT IAKLSGGEKKRLYLLK+LIEKPNVLLLEPTNDLDIATL VLENFL

Sbjct: 421 LEQFLFPRSSHGTIAKLSGGEKKRLYLLKILIEKPNVLLLEPTNDLDIATLKVLENFL 480

Query: 481 QGFGGFPVITVSHDRYFLDKVANKIIAFEDNDIREFFGNYYTDYLDEKAFNEQNNNEVSKKE 540
F GPVITVSHDRYFLDKVA KI+AFE+ DIR F+GNY+DYLDEK F ++ E K

-2192-

Sbjct: 481 ANFAGPVITVSHDRYFLDKVATKILAFEEGDIRVFGNYSDYLDEKVFETV EADLAKT 540

Query: 541 STKTS---REKQSRKRMSYFEKQEWATIEDDIMILENTYTRIENDMQTCGSDFTRLSDLQ 597
 + +K+ RKMSY EKQEW IED I +E I IEN M T SD+ +L+ LQ

5 Sbjct: 541 TVTEEVPLPQKEERKRMSYLEKQEWAIQIEDKIATIEANIEEENQMLTVVSDYGQLAQLQ 600

Query: 598 KELDAKNEALLEKYDRYEYLSGLD 621
 KELD +N LL Y+R+EYLS LD

10 Sbjct: 601 KELDQRNNDLLAYERFEYLSGLD 624

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1948

A DNA sequence (GBSx2057) was identified in *S.agalactiae* <SEQ ID 6035> which encodes the amino acid sequence <SEQ ID 6036>. This protein is predicted to be poly(a) polymerase (papS). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2658(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9957> which encodes amino acid sequence <SEQ ID 9958> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB38446 GB:L47709 poly(A) polymerase [Bacillus subtilis]
 Identities = 157/395 (39%), Positives = 235/395 (58%), Gaps = 14/395 (3%)

30 Query: 11 FQKALPILKKIKKAGYEAYFVGGSVRDVLLDRPIHDVDIATSSYPEETKQIFKRTVDVGI 70
 F KALP+L+ + +AG++AYFVGG+VRD + R I DVDIAT + P++ +++F+RTVDVG
 Sbjct: 5 FIKALPVLRIIEAGHQAYFVGGA VRDSYMKRTIGDVIDIATDAAPDQVERLFQRTVDVVGK 64

35 Query: 71 EHGTVLVLEKGGYEITTFRTTEEYVVDYRRPSQVNFVRSLEEDLKRRDFTVNAFALNEDG 130
 EHGT++VL + YE+TTFRTE YVD+RRPS+V F+ SLEEDLKRRD T+NA A+ DG
 Sbjct: 65 EHGTIIVLWEDETYEVTTFRTESDYVDFRRPSEVQFISSLEEDLKRRDLTINAMAMTADG 124

40 Query: 131 EVIDLFHGLDDLDNHLRAVGLASERFNEDALRIMRGLRFSASLNFDIETTTFEAMKKHA 190
 +V+D F G D+D ++R VG +RF EDALR++R +RF + L F + T EA+ K
 Sbjct: 125 KVLDFYFGGKKDIDQKVIRTVGKPEDRFQEDALRMLRAVRFMSQLGFTLSPETEEATAKEK 184

45 Query: 191 SLLEKISVERSFIEFDKLLLAPYWRKGMALIDSHAFNYLPCLKNRELQLSAFLSQDKD 250
 SLL +SVER IEF+KLL R+ + LI + + LP ++ L +S +
 Sbjct: 185 SLLSHVSVERKTIEFEKLLQGRASQALQTLIQTRLVEELPGFYHKRENL---ISTSEFP 241

50 Query: 251 FLFETS-EQAWASLILSMEV--EHTKTFLKKWKTSTHFQKDVEHIVDVYRIREQMGLTKE 307
 F TS E+ WA+L++++ + + FLK WK K+ HI D + L
 Sbjct: 242 FFLTSREELWAALLINLGIVLKDAPFLKAWKLPKGVIKEAIIHADTF---GQSLDAM 297

55 Query: 308 HLYRYGKTTIIKQAE GIRKAR-GLMVDFEKIEQLD---SELAIHDRHEIVVNGGTLIKKL 363
 +YR GK + A I + R +D +K++ + L I ++ + G L+
 Sbjct: 298 TMYRAGK KALLSAAKISQLRQNEKLD EKKLDI QYAYQNLPIKSLKDL DITGKDLLALRN 357

Query: 364 IKPGPQM GDIISQIELAIVLGQLINEEEAILHFVK 398
 G + + + IE A+V G+L N+++ I ++K
 Sbjct: 358 RPAGKWWSEELQWIEQAVVTGKLSNQKKHIEEWLK 392

-2193-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6037> which encodes the amino acid sequence <SEQ ID 6038>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2023(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 256/400 (64%), Positives = 312/400 (78%)

```

Query: 2  MRLNYLPSEFQKALPILKKIKKAGYEAYFVGGSVRDVLLDRPIHDVDIATSSYPEETKQI 61
          M+L +PSEFQKALPIL KIK+AGYEAYFVGGSVRDVLL+RPIHDVDIATSSYPEETK I
Sbjct: 1  MKLMTMPSEFQKALPILTKIKEAGYEAYFVGGSVRDVLLERPIHDVDIATSSYPEETKAI 60

Query: 62  FKRTVDVGIEHGTVLVLEKGGYEITTFRTEEVYVDYRRPSQVNFVRSLEEDLKRRDFTV 121
          F RTVDVGIEHGTVLVLE GGEYEITTFRTE++YVDYRRPSQV+FVRSLEEDLKRRDFTV
Sbjct: 61  FNRTVDVGIEHGTVLVLENGGEYEITTFRTEDIYDYRRPSQVSFVRSLEEDLKRRDFTV 120

Query: 122  NAFALNEDGEVIDLFHGLDDLNDHLLRAVGLASERFNEDALRIMRGLRFSASLNFDIETT 181
          NA AL+E+G+VID F GL DL LRAVG A ERF EDALRIMRG RF+ASL+FDIE
Sbjct: 121  NALALDENGQVIDKFRGLIDLKQKRLRAVGKAERFEEDALRIMRGFRFAASLDFDIEAI 180

Query: 182  TFEAMKKHASLLEKISVERSFIEDKLLLPYWRKGMLALIDSHAFNYLPCLKNRELQLS 241
          TFEAM+ H+ LLEKISVERSF EFDKLL+AP+WRKG+ A+I A++YLP LK +E L+
Sbjct: 181  TFEAMRSHSPILLEKISVERSFTEFDKLLMAPHWRKGISAMIAQAYDYLPGKQQEAGLN 240

Query: 242  AFLSQLDKDFLFETSEQAWASLILSMEVEHTKTFLKWKTSSTHFQKDVEHIVDVYRREQ 301
          + L +F F QAWA +++S+ +E K+FLK WKTS FQ+ V ++ +YRIR++
Sbjct: 241  HLIVSLKDNFTFSDYHQAWAYVMISLAIEDPKSFLKAWKTSNDFQRYVTKLIALYRIRQE 300

Query: 302  MGLTKEHLYRYGKTIKQAEGRKARGLMVDFEKIEQLDSELAHDRHEIVVNGGTLIKK 361
          K +Y+YGK + E +RKA+ L VD ++I LD L IHD+H+IV+NG LIK
Sbjct: 301  RSFEKLDIYQYGGKMASLVEDLRKAQSLSDMDRINTLDQALVIHDKHDIVLNGSHLIK 360

Query: 362  LGIKPGPQMGGDIISQIELAIVLGQLINEEEAILHFVKQYL 401
          G+K GPQ+G ++ ++ELAIV G+L N+ I FV++ L
Sbjct: 361  FGMKSGPQLGLMLEKVELAIVEGRLDNDFTTIEAFVREEL 400

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1949

A DNA sequence (GBSx2058) was identified in *S.agalactiae* <SEQ ID 6039> which encodes the amino acid sequence <SEQ ID 6040>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2939(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07346 GB:AP001519 unknown conserved protein [Bacillus halodurans]

Identities = 94/274 (34%), Positives = 153/274 (55%), Gaps = 2/274 (0%)

Query: 2 KLALITDT SAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLQYYDKLAASKELPKT 61

-2194-

K+A++TD++AYL V V+ + ++ + Y E L+ +Y+KL ++LP T
 Sbjct: 3 KIAIVTDSTAYLGPKRAKELGVIVVPLSVVFGEAYQEEVELSSADFYEKLKHEEKLPTT 62
 Query: 62 SQPSLAELDDLCLQLEKEGYTHVLGLFIAAGISGFWQNIQFLIEEHPNLTIAFPDTKIT 121
 SQP++ + +L KEG+ V+ + +++ ISG +Q+ + + D+ I+
 Sbjct: 63 SQPAVGLFVETFERLAKEGFEVVISIHLSSKISCTYQSALTAGSMVEGIEVIGYDSGISC 122
 Query: 122 APQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNLHLVKGGRLSNGSAIIG 181
 PQ N V A +EG D I++ + ++ VV+DL+HL +GGRL+ ++G
 Sbjct: 123 EPQANFVAEAAKLKVEGADPQTIIDHLDEVKKRTNALFVVDLHLHGGRLNAAQLVVG 182
 Query: 182 NLLSIKPVLFHFNEEGKIVVYEKVRTEKKALKRLAEI-VKEMTADGEYDIAIHSRAQDKA 240
 +LL IKP+LHF E+G IV EKVRTEKKA R+ E+ +E ++ +IH+ D A
 Sbjct: 183 SLLKIKPILHF-EDGSIVPLEKVRTEKKAWARVKELFAEEASSASSVKATVIHANRLDGA 241
 Query: 241 EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEAV 274
 E+L + + D+ I FG VI THLGE++
 Sbjct: 242 EKLADERSQFSHVVDVSIHFPGVIGTHLGEESI 275

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6041> which encodes the amino acid sequence <SEQ ID 6042>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3379(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 181/281 (64%), Positives = 233/281 (82%)

Query: 1 MKLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLQYYDKLAASKELPK 60
 MKLA+ITD++A LP ++ + ++ LDIP+IID +TY EG+NL++D +Y K+A S+ LPK
 35 Sbjct: 1 MKLAVITDSTATLPTDLKQDKAIFSLDIPVIIDDETYFEGRNLSIDDFYQKMADSQNLPK 60
 Query: 61 TSQPSLAELDDLCLQLEKEGYTHVLGLFIAAGISGFWQNIQFLIEEHPNLTIAFPDTKIT 120
 TSQPSL+ELD+LL L +GYTHV+GLF+A GISGFWQNIQFL EEHP + +AFPD+KIT
 Sbjct: 61 TSQPSLSELDNLLGLLSSKGYTHVIGLFLAGGISGFWQNIQFLAEHPPEIEMAFPSKIT 120
 Query: 121 SAPQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNLHLVKGGRLSNGSAII 180
 SAP G++V+N L SR+GM F I+NK+Q QI+ FI+V+DLNLHLVKGGRLSNGSA++
 Sbjct: 121 SAPLGSVMKNVLDWSRQGMTFQAILNKLQEIQIDGTTAFIMVDDLNLHLVKGGRLSNGSALL 180
 Query: 181 GNLLSIKPVLFHFNEEGKIVVYEKVRTEKKALKRLAEIVKEMTADGEYDIAIHSRAQDKA 240
 GNLLSIKP+L F+EEGKIVVYEKVRTEKKA+KRL EI+ ++ ADG+Y++ IHS+AQDKA
 Sbjct: 181 GNLLSIKPIRFDEEGKIVVYEKVRTEKKAMKRLVEILLNDLIADGQYNVFIHSKAQDKA 240
 Query: 241 EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEAVAFGITPK 281
 + L LL +G + D+E V FG VIATHLGEA+AFG+TP+
 50 Sbjct: 241 DYLRLLQDSGYQYDIEEVHFGAVIATHLGEAIAFGVTPR 281

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1950

A DNA sequence (GBSx2059) was identified in *S.agalactiae* <SEQ ID 6043> which encodes the amino acid sequence <SEQ ID 6044>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.59 Transmembrane 51 - 67 (50 - 67)

60

-2195-

----- Final Results -----

bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6045> which encodes the amino acid sequence <SEQ ID 6046>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.19 Transmembrane 50 - 66 (49 - 67)

----- Final Results -----

bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/126 (74%), Positives = 115/126 (90%)

Query: 1 MEVIREQEFVNQYHYDARNLEWEENGTPKTNFEVTFQLANRDEAAKVTISIVAVLQFVIV 60
 M+++RE+EFVNQYHYDARNLEWE+ENGTP+TNFEVTFQL ++DE K T IV+VLQFVIV
 Sbjct: 1 MQLVREKEEFVNQYHYDARNLEWEKENGTPETNFEVTFQLIDKDEQQKETVIVSVLQFVIV 60

Query: 61 RDEFVISGVISQMAHIQGRLINEPSEFSQDEVENLAAPLLEIVKRLTYEVTEIALDRPGV 120
 ++EFVISGVISQ I RL+++PSEF+Q+EVE+LAAPLL++VKRLTYEVTEIALDRPG+
 Sbjct: 61 KEEFVISGVISQMRILDRLDKPKSEFTQEEVESLAAPLLDMVKRLTYEVTEIALDRPGI 120

Query: 121 TLEFNS 126
 LEF +
 Sbjct: 121 HLEFKN 126

SEQ ID 6044 (GBS416) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 4; MW 17.5kDa).

GBS416-His was purified as shown in Figure 214, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1951

A DNA sequence (GBSx2060) was identified in *S.agalactiae* <SEQ ID 6047> which encodes the amino acid sequence <SEQ ID 6048>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3875(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2196-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1952

A DNA sequence (GBSx2061) was identified in *S.agalactiae* <SEQ ID 6049> which encodes the amino acid sequence <SEQ ID 6050>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1953

A DNA sequence (GBSx2062) was identified in *S.agalactiae* <SEQ ID 6051> which encodes the amino acid sequence <SEQ ID 6052>. This protein is predicted to be PTS system, fructose-specific enzyme II, BC component (fruA-1). Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

| | | | |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.56 | Transmembrane | 630 - 646 (618 - 653) |
| INTEGRAL | Likelihood = -7.43 | Transmembrane | 307 - 323 (303 - 331) |
| INTEGRAL | Likelihood = -7.01 | Transmembrane | 415 - 431 (412 - 435) |
| INTEGRAL | Likelihood = -7.01 | Transmembrane | 448 - 464 (444 - 474) |
| INTEGRAL | Likelihood = -3.72 | Transmembrane | 595 - 611 (591 - 612) |
| INTEGRAL | Likelihood = -3.61 | Transmembrane | 530 - 546 (529 - 553) |
| INTEGRAL | Likelihood = -2.39 | Transmembrane | 350 - 366 (350 - 371) |
| INTEGRAL | Likelihood = -1.70 | Transmembrane | 486 - 502 (486 - 506) |
| INTEGRAL | Likelihood = -1.49 | Transmembrane | 376 - 392 (376 - 392) |

----- Final Results -----

bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9959> which encodes amino acid sequence <SEQ ID 9960> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC component [Bacillus halodurans]

Identities = 320/659 (48%), Positives = 438/659 (65%), Gaps = 46/659 (6%)

| | |
|-----------|---|
| Query: 1 | MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDITGVVTFNFAIFKDGIMKREAQTSTGLG 60 |
| | +KI +LLKK+ M+++L+A SKEA IDE++ L G + + FK I++RE+Q++TG+G |
| Sbjct: 2 | LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61 |
| Query: 61 | DGIAMPHSKNAAVKEATVLFASASGVDYEDLDGQPTDLFFMIAAPDGANDTHLALAE 120 |
| | +GIA+PH+K AAVK+ + F +S +G+DYE+LDGQP+ LFFMIAA +GAN+ HL L+ L |
| Sbjct: 62 | EGIAIPHAKTAAVKQPAIAFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETL 121 |

-2197-

Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
S +L+ E F L +A++ D+I+A D +E + +G + ++AVT C TG
Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAID---KKEAETAGEAEKQEGYE-LLAVTGCPTG 176

Query: 181 IAHTYMAEEALKKKAAEMGVGKIVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
IAHTYMA + LK KA E+GV IKVETNG+ GV N+LT +I+ AK +I+AAD VEMDRF
Sbjct: 177 IAHTYMAADNLKSKAQELGVSIVKIVETNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRF 236

Query: 241 DGKPLVSRPVADGIKKSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSGLS---AFYK 297
GKP++ PV DGI++ ++LI+ L KA Y + Q+ DG +G G FYK
Sbjct: 237 HGKPVIVQVPVIDGIRRPKELIDQALACKAPVY----EGGAQASGEDGSAGGGRPKLG FYK 292

Query: 298 HLMGGVSQMLPFVIGGGIMIAIAFLDNILGVPKDQSLNLSYHEIAALFKNIGGA-AFA 356
HLM GVS MLPFV+GGGI+IAI+F+F P D SYH A + IGG AF
Sbjct: 293 HLMNGVSNMLPFVVGGLIAISFMFGIKAFDPSPD-----SYHPFAEMLMTIGGNAFG 347

Query: 357 FMLPVLGYIAYSIAEKPLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLG 416
M+PVLA +IA SIA++PG AG + G IAS+G A GFLG
Sbjct: 348 LMIPVLAFAFIAMSIADRPFGFAAGMIGGLIASTGEA-----GFLG 386

Query: 417 ALVGGFLAGGVILLRKLKLSGLPKSLEGIKSILLYPLLGLVITGFLMLLVNIPMAAINTA 476
L+ GFLAG V L ++K+L+ LP++L+GIK+IL YP+ + ITG +ML++ P+AA NT
Sbjct: 387 GLIAGFLAGYVALGVKKVLANLPQTLTGKILFYFVFNIFITGMIMLVIVGPLAAFNTG 446

Query: 477 LNTFLQGLSGSSAVLMGLLVGGMAVDMGGPKNKAAVFGTGTLAATVANGGSSVMAAVM 536
L +L + ++ V++G+++GGMAVDMGGP+NKAA+ FG + A G AAVM
Sbjct: 447 LQDWLGSMTANMVILGVILGGMMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAVM 502

Query: 537 AGGMVPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSI 596
AGGMVPPL + +AT LFK KF +ER++G TN ++G SFITEGAIPF AADP R IPS I
Sbjct: 503 AGGMVPLGLIALATTLFKKKFTKQEREAGKTNVILGASFITEGAIPFAAADPGRVIPSI 562

Query: 597 VGSALTGALVGLAGIKLMAHPGGIFVI---ALTSNPLLYILFILIGAVVSGVLFGLFRK 652
VGSA G L L + L APHGG FVI + +NPLLY++ I+ G++V+ +L G ++K
Sbjct: 563 VGSAGGLTALFNVTLSPHGGAFVIFIGNIVNPLLYLVAILIAGSIVTALLLGFWKK 621

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6053> which encodes the amino acid sequence <SEQ ID 6054>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

| | | | |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.77 | Transmembrane | 624 - 640 (612 - 646) |
| INTEGRAL | Likelihood = -7.59 | Transmembrane | 301 - 317 (297 - 321) |
| INTEGRAL | Likelihood = -6.85 | Transmembrane | 442 - 458 (439 - 468) |
| INTEGRAL | Likelihood = -5.95 | Transmembrane | 409 - 425 (406 - 426) |
| INTEGRAL | Likelihood = -3.61 | Transmembrane | 524 - 540 (523 - 547) |
| INTEGRAL | Likelihood = -2.50 | Transmembrane | 337 - 353 (337 - 353) |
| INTEGRAL | Likelihood = -2.44 | Transmembrane | 589 - 605 (589 - 605) |
| INTEGRAL | Likelihood = -1.70 | Transmembrane | 480 - 496 (480 - 500) |
| INTEGRAL | Likelihood = -1.44 | Transmembrane | 370 - 386 (370 - 386) |

----- Final Results -----
bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC component [Bacillus halodurans]
Identities = 322/659 (48%), Positives = 431/659 (64%), Gaps = 48/659 (7%)

Query: 1 MKIQDLLRKDIMILDQAISKEVAIDEMITKLVKDIVHDFDVFKKSIMTREEQTSTGLG 60
+KI +LL+KD M+L+L+A SKE IDE++ L + ++D FK++I+ RE Q++TG+G
Sbjct: 2 LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61

-2198-

Query: 61 DGIAMPHSKNIVVDKPAVLFAKSNKGVVDYKALDGOPTDLFFMIAAPQGANDTHLAALAE 120
+GIA+PH+K V +PA+ F +S+ G+DY++LDGQP+ LFFMIAA +GAN+ HL L+ L
Sbjct: 62 EGIAlPHAKTA AVKQPAI AFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETLSRL 121

5 Query: 121 SQYLLKDG FADKLRAAATPEAVIAVFD--EASTAKEEVVAPTSGQDFIVAVTACPTGIAH 178
S +L+ + F L A + + +A D EA TA E + ++AVT CPTGIAH
Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAIDKKEAETAGEAEBEKQEGYE--LLAVTGCPTGIAH 179

10 Query: 179 TYMAEEALKKQAAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRFDGK 238
TYMA + LK +A E+GV+IKVETNG+ GV NRIT E+I AK +IVAAD VEMDRF GK
Sbjct: 180 TYMAADNLKSKAQELGVSIKVETNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRPHGK 239

15 Query: 239 QFIARPVADGIKKSQELISLILNNEGNTYHAKNGKSETAVSTEKTSLGG-----AFYKHL 293
I PV DGI++ +ELI L + Y + S E S GG FYKHL
Sbjct: 240 PVIQVPVTDGIRPKELIDQALAGKAPVY-----EGGAQASGEDGSAGGGRPKLGFYKHL 294

20 Query: 294 MGGVSQMLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIGGA-AFSFM 352
M GVS MLFPV+GGGI+IA++F+ P+D SYH A + M IGG AF M
Sbjct: 295 MNGVSNMLPFVVGGLIAISFMFGIKAFDPSPD-----SYHPFAEMLMTIGGNAFGLM 349

25 Query: 353 LPVLAGYIAYSIAEKPLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPSGFLGAL 412
+PVIA +IA SIA++PG AG + G IAS G A GFLG L
Sbjct: 350 IPVLAAFIAMSIAIDRPGFAAGMIGGLIASTGEA-----GFLGGL 388

30 Query: 413 VGGFLAGGVILALRKLLAGLPRSLLEGVKSILLYPLLGLVLTGFLMLFVNIPMAAINTALN 472
+ GFLAG V L ++K+LA LP++L+G+K+IL YP+ + +TG +ML + P+AA NT L
Sbjct: 389 IAGFLAGYVALGVKKVLANLPQTLTGKILFYFPVFNIFITGMIMLVIVGPLAAFNTGLQ 448

35 Query: 473 DFLQGLSGSSAVLMGLLVGGMAVDMGGPVNKAAYVFGTGTAAATVANGGSVMAAVMAG 532
D+L + ++ V++G+++GGMAVDMGGP+NKAA+ FG + A G AAVMAG
Sbjct: 449 DWLGSMGTANMVILGVILGGMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAMVAG 504

40 Query: 533 GMVPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSFIA 592
GMVPPL + +AT LFK KFTK+ERE+G TN ++G SFITEGAIPF AADP R IPS I G
Sbjct: 505 GMVPPLGIALATTLFKKFTKQEREAGKTNILGASFITEGAIPFAAADPGRVIPSIIVG 564

Query: 593 SALTGALVGLAGIKLMAHPGGIFVI---ALTSNPILYLVFVVGALVSGILFGALRKA 648
SA G L L + L APHGG FVI + +NP+LYLV ++ G++V+ +L G +K A
Sbjct: 565 SAFAGGLTALFNVTL SAPHGGAFVIFIGNIVNNPLLYLVAIAGSIVTALLLGFWKDA 623

An alignment of the GAS and GBS proteins is shown below.

Identities = 526/652 (80%), Positives = 581/652 (88%), Gaps = 6/652 (0%)

45 Query: 1 MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDITGVVTNFAIFKDGIMKREAQTSTGLG 60
MKIQDLL+K++MI+DL+A SKE AIDEMITKL+ +V +F +FK IM RE QTSTGLG
Sbjct: 1 MKIQDLLRKDIMILDQAISKEVAIDEMITKLVEKDIVHDFDVFKKSIMTREEQTSTGLG 60

50 Query: 61 DGIAMPHSKNAVKEATVLFASASGVVDYALDGOPTDLFFMIAAPDGANDTHLAALAE 120
DGIAMPHSKN V + VLFAS GVDY+ALDGOPTDLFFMIAAP GANDTHLAALAE
Sbjct: 61 DGIAMPHSKNIVVDKPAVLFAKSNKGVVDYKALDGOPTDLFFMIAAPQGANDTHLAALAE 120

55 Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
S+YLLK+GFAD+LR A TP+ +IA FD S ++E V T G D+IVAVTAC TG
Sbjct: 121 SQYLLKDG FADKLRAAATPEAVIAVFD--EASTAKEEVVAPT-----SQQDFIVAVTACPTG 175

60 Query: 181 IAHTYMAEEALKKKAEMGVGKIKVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
IAHTYMAEEALKK+AAEMGV IKVETNGASGV N+LT+ DI RAKGVI+AADKAVEMDRF
Sbjct: 176 IAHTYMAEEALKKQAAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRF 235

65 Query: 241 DGKPLVSRPVADGIKKSEDLINILDNKAQTYHAKNQNDKQSGESDGKSLGSAFYKHL 300
DGK ++RPVADGIKK++LI++IL+N+ TYHAKN ++ S K+ LG AFYKHL
Sbjct: 236 DGKQFIARPVADGIKKSQELISLILNNEGNTYHAKN-GKSETAVSTEKTSLGGAFYKHL 294

Query: 301 GGVSQMLPFVIGGGIMIAAFLFDNMLGVPNDQLGSLGSYHEIAALFKNIGGAFAFM 360
GGVSQMLPFVIGGGIMIA+AF L DN+LGVP DQL +LGSYHEIAA+F NIGGAFAFM
Sbjct: 295 GGVSQMLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIGGAFAFM 354

-2199-

Query: 361 VLAGYIAYSIAEKPGLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLGALVG 420
 VLAGYIAYSIAEKPGLVAGFVAG+IAS+GLAFGKVPFA GG+ +L L GVPSPGFLGALVG
 Sbjct: 355 VLAGYIAYSIAEKPGLVAGFVAGAIASNGLAFLGKVPFAAGGEVSLGLTGVPSPGFLGALVG 414

5 Query: 421 GFLAGGVILLRKLKLLSGLPKSLEGIKSILLYPLLGVLTGFLMLLVNIPMAAINTALNTF 480
 GFLAGGVIL LRKLL+CLP+SLEG+KSILLYPLLGVLT+TGFLML VNIPMAAINTALN F
 Sbjct: 415 GFLAGGVILALRKLKLAGLPRSLGKVSILLYPLLGVLTGFLMLFVNIPMAAINTALNDF 474

10 Query: 481 LQGLSGSSAVLMGLLVGGMAVDMGGPVPNKAAYVFGTGTLAATVANGGSVMAAVMAGGM 540
 LQGLSGSSAVLMGLLVGGMAVDMGGPVPNKAAYVFGTGTLAATVANGGSVMAAVMAGGM
 Sbjct: 475 LQGLSGSSAVLMGLLVGGMAVDMGGPVPNKAAYVFGTGTLAATVANGGSVMAAVMAGGM 534

15 Query: 541 VPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSFIVGSA 600
 VPPLAVFVATLLFKDKF EER+SGLTNIVMGLSFITEGAIPFGAADPARAIPSF I GSA
 Sbjct: 535 VPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSF IAGSA 594

Query: 601 LTGALVGLAGIKLMAHPGGIFVIALTSNPLLYILFILIGAVVSGVLFGLFRK 652
 LTGALVGLAGIKLMAHPGGIFVIALTSNP+LY++F++IGA+VSG+LFG RK
 Sbjct: 595 LTGALVGLAGIKLMAHPGGIFVIALTSNPILYLVFVVIGALVSGILFGALRK 646

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1954

A DNA sequence (GBSx2063) was identified in *S.agalactiae* <SEQ ID 6055> which encodes the amino acid sequence <SEQ ID 6056>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1532(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC24914 GB:AF012285 fructose-1-phosphate kinase [Bacillus subtilis]
 Identities = 146/303 (48%), Positives = 197/303 (64%)

Query: 1 MIYTVTLNPSIDFIVRLDTLLLGSVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
 MIYTVTLNPS+D+IV ++ +G +NR + D KY GGKGINVSR+LKR + + A GF+G
 40 Sbjct: 1 MIYTVTLNPSVDYIVHVEDFTVGGLNRSSYDTKYPGGKGINVSRLLKRHHVASKALGFVG 60

Query: 61 GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVVKAKIETEINGGGPRITNEQLHRLEKLL 120
 GFTG +++ L E ++T F V DTRINVK+K ETEING GP I++E +
 45 Sbjct: 61 GFTGEYIKTFLREENLETAFFSEVKGDTTRINVKLKTGDETEINGQGPTISDEDFKAFLEQF 120

Query: 121 SRLTPEDTVVFAGSAPASLGNKYNTLIPIAKKTGAEEVVCDFEGQTLILDALAYQPLLVKP 180
 L D VV AGS P+SL + Y + K+ A VV D G+ LL A +P L+KP
 Sbjct: 121 QSLQEGDIVVLGSIPLSPHDTYEKIAEACKQONARVVLDISGEALLKATEMKPFLMKP 180

50 Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
 N+HEL ++FG + + + Y K++++GA++VIVSMAGDGALL T EA YFA KG+
 Sbjct: 181 NNHELGMFGTAITSVEEAVPYGKLVQGAEHVIVSMAGDGALLFTNEAVYFANVPK GK 240

55 Query: 241 VKNSVGAGDSMVAGFTGEFVSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
 + NSVGAGDS+VAGF K EA + GV G+AT FS++L T EF+Q + +V+V
 Sbjct: 241 LVNSVGAGDSVVAGFLAGISKQLPLEEAFRLGVTSGSATAFSEELGTEEFVQQLLPEVKV 300

Query: 301 EKL 303
 +L
 60 Sbjct: 301 TRL 303

-2200-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6057> which encodes the amino acid sequence <SEQ ID 6058>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 222/302 (73%), Positives = 261/302 (85%)

```

Query: 1  MIYTVTLNPSIDFIVRLDTLLGSVNRMSTDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
          MIYTVTLNPSIDFIVR+D + LGSVNRM SDDK+ GGKGINVSRIL+RL I +TATGF+G
Sbjct: 1  MIYTVTLNPSIDFIVRIDQINLGSVNRMASDDKFAGGKGINVSRILQRLDIASATGFLG 60

Query: 61  GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVVKVAKIETEINGGGPRITNEQLHRLEKLL 120
          GFTG F+E+ L EG+KTDFV ++DTRINVK+K++ ETE+NG GP I+ EQL L+ L
Sbjct: 61  GFTGRFIEESLSAEGVKTD FVKGDQDTRINVKIKSQEETELNGQGPIISQELEDLKTCL 120

Query: 121 SRLTPEDTVVFAGSAPASLGNKVYNTLIPIAKKTGAEVVCD FEGQTL LDALAYQPLL VKP 180
          S+LT EDTVVFAGSAPA+LGN VY L+P+ +++GA+VVCDFEGQTL+DALAY PLLVKP
Sbjct: 121 SQLTAEDTVVFAGSAPANLGNNAVYKELLPLVRQSGAQVVCDFEGQTLIDALAYNPLL VKP 180

Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
          NNHEL IFG L L D+E YA ++L+ GA+NVI+SMAGDGALLVT EA+YFAKPIKGE
Sbjct: 181 NNHELEAIFGTILTSLDDVETYARRLLEMGQNVIISMAGDGALLVTKEATYFAKPIKGE 240

Query: 241 VKNSVGAGDSMVAGFTGEFVKSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
          VKNSVGAGDSMVAGFTGEF+KS+NP+EALKWGVACGTAT FSDDLAT FI++ Y+KVEV
Sbjct: 241 VKNSVGAGDSMVAGFTGEFMKSONPIEALKWGVACGTATATFSDDLATIAFIKETYHKVEV 300

Query: 301 EK 302
          EK
Sbjct: 301 EK 302

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1955

A DNA sequence (GBSx2064) was identified in *S.agalactiae* <SEQ ID 6059> which encodes the amino acid sequence <SEQ ID 6060>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9961> which encodes amino acid sequence <SEQ ID 9962> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24913 GB:AF012285 FruR [Bacillus subtilis]

Identities = 97/247 (39%), Positives = 148/247 (59%), Gaps = 4/247 (1%)

Query: 23 MLKSKRKEIILSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGALPY 82

-2201-

ML +R ++I+ ++E++ V + EL ++ SEST+RRDL LE GFLKRVHGG
 Sbjct: 1 MLTPERHQLIIDQIEKHDDVVKIQELINLTNASESTIRRDLSLEERGFLKRVHGGAAKLS 60

Query: 83 SLGQELSNQEKAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPH-EQLTVV 141
 + E EK+ KN+ KL IA + A L+ + D I++DAGTTT +IDF+ + + VV

Sbjct: 61 DIRLEPDMLEKSSKNLHDKLKIAEKAASLLEEGDCIYLDAGTTTLHMIDFMDKTKDIVVV 120

Query: 142 TNSIHHAAKLVDRGIKTIIGGAVKHSTDASIGQVAINQIRQITVDKAFLGMNGID-EVY 200
 TN + H L+ + I ++GG VKH T A IG ++ + Q DK+FLG NG+ E

Sbjct: 121 TNGVMHIDALIRKEISFYLLGGYVKHRTGAIIGGASLVAMDQYRFDKSFLGTNGVHTEAG 180

Query: 201 LTPDLEEAIAKEAIINNSQQTFILMDSSKIGQVTFKVKKEINDINLVTKTDSSELMTII 260
 TTPD +EA +K+ I ++ ++L D SK G+++F+ I D ++T TD+E +T

Sbjct: 181 FTTPDPDEALLKQKAIKQAKHAYVLADPSKFGEISFSAFAGIGDATIIT--TDAEELTFD 238

Query: 261 KEKMKVI 267
 + K +

Sbjct: 239 NYQEKTV 245

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6061> which encodes the amino acid sequence <SEQ ID 6062>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2604(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 135/237 (56%), Positives = 184/237 (76%)

Query: 33 LSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGAEPLPYSLGQELSNQE 92
 ++++ + V+L++L +L +SEST+RRDL ELE G L RVHGGAE L +SL +ELSNQE

35 Sbjct: 1 MAKITEENYVSLDLMLQLLNSSESTIRRDLEGELEGEGRHLRVHGGAE L FHS LQEELSNQE 60

Query: 93 KAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPHEQLTVVTNSIHHAALKV 152
 K++KN K IA++ ++LI DVIFIDAGTTTE L+ FL + LTVVTNSIHHAALV

40 Sbjct: 61 KSVKNSHIKKAIQASQLIYDNDVIFIDAGTTTEFLPFLQAKNLTVVTNSIHHAARLV 120

Query: 153 DRGIKTIIGGAVKHSTDASIGQVAINQIRQITVDKAFLGMNGIDEVYLTTPDLEEAIAK 212
 + I+TII+GG VK +TDASIG VA+ QIRQ+ DKAFLGMNG+D+ YLTTPD+EEA IK

Sbjct: 121 ELSIETIIVGGYVKQTTDASIGNVALEQIRQMNFDKAFLGMNGVDDSYLTTPDMEEAVIK 180

45 Query: 213 EAIINNSQQTFILMDSSKIGQVTFKVKKEINDINLVTKTDSSELMTIIKEKMKVIQV 269
 +A+++N++ +IL+D +KIGQV+F KV IND+ ++T + ++ IKEK KVI++

Sbjct: 181 KAVLSNAKLAYILVDGKIGQVSFVKVAPINDVTIITLGGASILKQIKEKAKVIEL 237

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1956

A DNA sequence (GBSx2065) was identified in *S.agalactiae* <SEQ ID 6063> which encodes the amino acid sequence <SEQ ID 6064>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

55 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5777(Affirmative) < succ>

-2202-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB89121 GB:AJ277485 beta-lactam resistance factor
 [Streptococcus pneumoniae]
 Identities = 215/410 (52%), Positives = 283/410 (68%)

10 Query: 1 MTLRELTIIEEFKEHSGNYDSQSFLQTPEMAKLLEKRGYDVRYLGYQVENKLEIISLSYIM 60
 M L LT EEF+ +S S+SF+Q+ +M LLEKRG + YL + E ++++ +L Y +
 Sbjct: 1 MALTLTLTKEEFQTVSDQVSSRSFMSQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60

15 Query: 61 PVTGGFQMKIDSGPVHSNSKYLKQFYKALQGYAKSNGVLELIVEPYDDYQLFTSSGVPSPN 120
 P+ GG M+++SGP+++ L FY L+ YAK NGVLEL+V+PY+ YQ F S G P +
 Sbjct: 61 PMLGGLHMEINSGPIYTQDALPVPFYAELKEYAKQNGVLELLVVKPYETYQTFDSQGNPID 120

20 Query: 121 QGNDNLIEDFTSSGYHHDGLTTGFTGKYLSWHYVKNLEGVTSSETLLSSFSKTGRALVKKA 180
 ++I+D T GY DGLTTG+ G W Y K+L +T ++LL SFSK G+ LVKKA
 Sbjct: 121 AEKKSIIQDLTDLGYQFDGLTTGYPGGEPDWLYYKDLTELTEKSLLKSFSKKGKPLVKKA 180

25 Query: 181 MSFGIKVRVLKRDELHLFKEITTSTSNRRDYMDKSLDYYQDFYDSFEGKAEFVIATLNFR 240
 +FGI+++ LKR+EL +FK IT TS RR+Y DKSL+YY+ FYD+F +AEF+IA+LNF
 Sbjct: 181 ETFGIRLKKLKREELSIFKNITKETSERREYSKSLLEYEHFYDTFGEQAEFLIASLNFS 240

30 Query: 241 EYDHNLIQIKAEALENKLKLLDERFRENADSPKYHRQRSEIINQLASFETRROEVQSFIQK 300
 +Y LQ + LE L L +N S K Q E +Q +FE R+ E + I+K
 Sbjct: 241 DYMSKLQGEQSKLEENLDKLRDLKSNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300

35 Query: 301 YDNQDVVLGSLFVYSLKETVYFFSGSYTEFNKFYAPAVLQEQYVMQEALKRGSTFYNLLG 360
 Y +D+VLGSLFVY +ET Y FSGSYTEFNKFYAPA+LQ+YVM E++KRG YN LG
 Sbjct: 301 YGEEDIVLAGSLFVYMPQETTYLFGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLG 360

Query: 361 IQGTFDGSDSLRFKQNFNGCIIRKMGTFTNYPSPFKYKGIQLLKKVLKR 410
 IQG FDGSD +LRFKQNFNG I+RK GTF Y+PSP KYK IQLLKK++ R
 Sbjct: 361 IQGIFDGS DGVLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410

There is also homology to SEQ ID 5460.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1957

A DNA sequence (GBSx2066) was identified in *S.agalactiae* <SEQ ID 6065> which encodes the amino acid sequence <SEQ ID 6066>. This protein is predicted to be cell wall protein, 40 kDa (sr 5' region). Analysis of this protein sequence reveals the following:

45 Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.45 Transmembrane 25 - 41 (23 - 42)

----- Final Results -----
 50 bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9963> which encodes amino acid sequence <SEQ ID 9964> was also identified.

55 The protein has homology with the following sequences in the GENPEPT database.

!GB:AF278686 choline binding protein D; CbpD [Strept...
 !GB:AF278686 choline binding protein D; CbpD [Strept...

-2203-

>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD

[Streptococcus pneumoniae]

Identities = 63/230 (27%), Positives = 108/230 (46%), Gaps = 34/230 (14%)

Query: 324 WTEQGGQDDIKWYTAVTTGDG-----NYKVAVSFADHKNEKGLYNIHLYQEASGTLVG 377
 W+ G + W + V GD NY S+ + +++++ G VG
 Sbjct: 123 WSTAGTYGHVAWVSVM-GDQIEIEEYNYGYTESYNKRVIKANTMTGFIHFKDLDDGGSVG 181

Query: 378 VTGTKVTVAGTNSSQEPIENGLAKTGVYNIIGSTEVEKNEAKISSQTQFTLEKGDKINYDQ 437
 + + + GT+ + + +K E S G+K++YDQ
 Sbjct: 182 NSQSTSTGGTHYFKT-----KSAIKTEPLASGTVIDYYYPGKVVHYDQ 225

Query: 438 VLTADGYQWISYKSYSGVRRYIPVKKLTTSSEKAKDEATKPTSYPNLPKTG-TYFTTKTV 496
 +L DGY+W+SY +Y+G RY+ ++ + + P L TG T+ F
 Sbjct: 226 LLEKDGKWLSTAYNGSYRYVQLEAVNKN-----PLGNSVLSSTGGTHYFKTKS 275

Query: 497 DVKSQPKVSSPVEFNFQKGEKIHYDQVLVVDGHQWISYKSYSGIRRYIEI 546
 +K++P VS+ V + GEK+HYDQ+L DG++W+SY +Y+G RRYI++
 Sbjct: 276 AIKTEPLVSATVIDYYYPGKVVHYDQILEKDGKWLSTAYNGSRRYIQL 325
 Identities = 49/161 (30%), Positives = 85/161 (52%), Gaps = 14/161 (8%)

Query: 116 GNYVYSKETEVKNTPSKSAPVAFYAKKGDVFDQVFNKDNVWISYKSFVCGVRRYAAIE 175
 G + + ++ +K P S V Y G+KV YDQ+ KD KW+SY ++ G RY +E
 Sbjct: 191 GTHYFKTKSAIKTEPLASGTVIDYYYPGKVVHYDQILEKDGKWLSTAYNGSRYVQLE 250

Query: 176 SLDPSCGGSETKAPTPTNSGSNNQEKIATQGNITFSHKVEVKNEAKVASPTQFTLDKGRD 235
 +++ + P+ NS + +T G + F K +K E V++ G++
 Sbjct: 251 AVNKN-----PLGNSVLS-----STGGTHYFKTKSAIKTEPLVSATVIDYYYPGEK 296

Query: 236 IFYDQILIEGNQWLSYKSFNGVRRFVLLGKASSVEKTEDK 276
 + YDQIL +G +WLSY ++NG RR++ L +S + +++
 Sbjct: 297 VHYDQILEKDGKWLSTAYNGSRRYIQLLEGVTSSQNYQNG 337
 Identities = 52/192 (27%), Positives = 90/192 (46%), Gaps = 13/192 (6%)

Query: 295 ISNETTIGFDILITNIKDDNGIAAVKVPVWTEQGGQDDIKWYTAVTTGDGNYKVAVSFAD 354
 I T TGF + KD +G + T GG K +A+ T + +
 Sbjct: 161 IKANTMTGF----IHFKDLDDGGSVGNSSQSTSTGGTHYFKTKSAIKTEPLASGTVIDYY- 215

Query: 355 HKNEKGLYNIHLY---YQEASGTLVGVTGKTAVAGTNSSQEPIENGLAKT--GVYNIIG 409
 + EK Y+ L Y+ S T + V + N + P+ N + + G +
 Sbjct: 216 YPGKVVHYDQILEKDGKWLSTAYNGSRYVQLEAVNKN--PLGNSVLSSTGGTHYFKT 273

Query: 410 STEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPVKKLTTSSE 469
 + +K E +S+ G+K++YDQ+L DGY+W+SY +Y+G RRYI ++ + TSS+
 Sbjct: 274 KSAIKTEPLVSATVIDYYYPGKVVHYDQILEKDGKWLSTAYNGSRRYIQLLEGV-TSSQ 332

Query: 470 KAKDEATKPTSY 481
 +++++ +SY
 Sbjct: 333 NYQNQSGNISSY 344
 Identities = 33/113 (29%), Positives = 56/113 (49%), Gaps = 2/113 (1%)

Query: 91 NTATKIDITPLVETKPMVEKTLPEQGNVYYSK-ETEVKNTPSKSAPVAFYAKKGDVFDY 149
 N + + + V P+ L G Y K ++ +K P SA V Y G+KV YD
 Sbjct: 241 NGSYRYVQLEAVNKNPLGNSVLSSTGGTHYFKTKSAIKTEPLVSATVIDYYYPGKVVHYD 300

Query: 150 QVFNKDNVWISYKSFVCGVRRYAAIESLDPSCGGSETKAPTPTNSGSNNQEKI 202
 Q+ KD KW+SY ++ G RRY +E + S + ++ +++ GS++ +
 Sbjct: 301 QILEKDGKWLSTAYNGSRRYIQLLEGVTSSQNYQNQSGN-ISSYGSSSSTV 352

A related GBS gene <SEQ ID 8937> and protein <SEQ ID 8938> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -6.74
 GvH: Signal Score (-7.5): 1.26
 Possible site: 42

-2204-

```
>>> Seems to have no N-terminal signal sequence
ALOM program      count: 1 value: -3.45 threshold: 0.0
  INTEGRAL        Likelihood = -3.45 Transmembrane 22 - 39 ( 23 - 42)
  PERIPHERAL      Likelihood = 6.26 371
modified ALOM score: 1.19
```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

The protein has homology with the following sequences in the databases:

41.2/57.9% over 283aa

Streptococ

cus mutans

EGAD|33594| cell wall protein, 40 kDa (sr 5' region) Insert characterized
PIR|A60328|A60328 40K cell wall protein precursor (sr 5' region) - (strain OMZ
175, serotype f) Insert characterized

ORF02145 (301 - 1803 of 2238)

EGAD|33594|34911(30 - 313 of 335) cell wall protein, 40 kDa (sr 5' region) {Streptococcus mutans|PIR|A60328|A60328 40K cell wall protein precursor (sr 5' region)} - Streptococcus mutans (strain OMZ175, serotype f)

```
%Match = 8.0
```

```
%Identity = 41.1    %Similarity = 57.9
```

Matches = 81 Mismatches = 79 Conservative Sub.s = 33

156 186 216 246 276 306 336 366
 *YA***FCYTKNKS~~WVFF~~SRSIYSIKYYICITNISKIC*HVTKRIL***CK*IRK*VFMMKKGVNDTKQSYSLRKYK
 : : : : | :| : : | | : |: ||
 MNQIVVISSFYMLGAHSFSKAVYHNDRSVKLMKRIDINHQAQRFSIRKYA
 10 20 30 40 50

396 426 456 486 516 546 576 606
FGLASVILGSFIMVTSVPFADQTTSVQVNNQTGTSVDANSSNETSASSVITSNNDSVQASDKVNSQNTATKDITPLV
|||::| : : : | : | :: : : || | : | ||
FGAASVLIGCVFFLGTQNVSAEQGTQL-----PASENAVNVNAENSVAISQAVADKAATQTTLTETPOV
70 80 90 100 110

654 684 714
ETKPMVEK-----TLPEQGNVYVYSKETEVKNTPSKSAIPVAF
| : | : : | | | | : | | | | : |
EVEEKESKVNAPALNVDDKGAKSKEDVN----AEQNEKAVRENLMCRQAKAVSIPSGQNVVFQETTPVKNAAAMSSP----

744 1533 1563 1593 1623 1653 1683
YAKKGDKVFDQVFNKD~~~GVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQVLTDGYQWISYKSYSGVRRYPV
|||::|||: || || |||:|||| |||:||| :
-----~-----TQFNFDKGDKVFDNVLEADGHQWISYVSYSGIRRYAPI
250 260 270

1713 1743 1773 1803 1833 1863 1893 1923
 KKLTTSSSEKAKDEATKPTSYPNLPKGTYYTFTKTVDVKSQPKVSSPVEFNQKGEKIHYDQVLVDGHQWISYKSYSGIR
 : :: | ||| ||| ||| :| : | :
 -----AVTIEELKQKEIVQONLPAQGTYHFTKQOQLKMLNCLVLRPNRSRFTTEITFFMIRF
 290 300 310 320 330

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6067> which encodes the amino acid sequence <SEQ ID 6068>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have a cleavable N-term signal seq.

-2205-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
 [Streptococcus pneumoniae]
 Identities = 93/217 (42%), Positives = 136/217 (61%), Gaps = 18/217 (8%)

Query: 42 GDNYPSPKWKKGNG-IDSWNMYIRQCTSF AAFRLSSANGFQLPKGYGNACTWGHIAKNQGY 100
 GD+YP+ +K G+ ID W MY RQCTSF AFRLS+ NGF++P YGNA WGH A+ +GY
 Sbjct: 51 GDDYPAYYKNGSQEIDQWRMYSRQCTSFVAFRLSNVNGFEIPAAYGNANEWGHRRAREGY 110

Query: 101 PVNKTPSIGAIWFDKNAYQSNAAAYGHVAVVADIRGDTVTIEEYNYNAGQGPERYHKRQI 160
 V+ TP+IG+I W + YGHVAVV+++ GD + IEEYNY E Y+KR I
 Sbjct: 111 RVDNTPITIGSITW-----STAGTYGHVAVVSNVMGDQIEIEEYNYGY---TESYNKRVI 161

Query: 161 PKSQVSGYIHFKDLSSQTSYPRQLKHISQASFDPSGT YHFTTRLPVKGQTSIDSPDLA 220
 + ++G+IHFKDL + + SQ+S GT++F T+ +K + +
 Sbjct: 162 KANTMTGFIHFKDLGGSGVGN-----SQSSTSTGGTHYFKTKSAIKTEPLASGTVID 213

Query: 221 YYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPIK 257
 YY G+ V+YD+++ GY WLSY +++G+ RY+ ++
 Sbjct: 214 YYYPGEKVHYDQILEKDGKWLSYTAYNGSYRYVQLE 250

An alignment of the GAS and GBS proteins is shown below.

Identities = 34/94 (36%), Positives = 52/94 (55%)

Query: 453 SGVRRYIPVKKLTTSSSEKAKDEATKPTSYPNLPKTGT YTFKTVDVKSQPKVSSPVEFNF 512
 S V YI K L++ + + K S + +GT Y FT + VK Q + SP +
 Sbjct: 163 SQVSGYIHFKDLSSQTSYPRQLKHISQASFDPSGT YHFTTRLPVKGQTSIDSPDLAYY 222

Query: 513 QKGEKIHYDQVLVVDGHQWISYKSYSGIRRYIEI 546
 + G+ ++YD+V+ G+ W+SY S+SG RRYI I
 Sbjct: 223 EAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256
 Identities = 30/78 (38%), Positives = 45/78 (57%), Gaps = 2/78 (2%)

Query: 402 TGVYNIIGSTEVKNEAKISSQTOFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPV 461
 +G Y+ VK + I S E G + YD+V+TA GY W+SY S+SG RRYIP+
 Sbjct: 197 SGT YHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256

Query: 462 KKLTTSSSEKAKDEATKPT 479
 K+ + +++ TKP+
 Sbjct: 257 KE--PAQSVVQNDNTKPS 272
 Identities = 27/94 (28%), Positives = 47/94 (49%)

Query: 198 NQEKIATQGN YTFSHKVEVKNEAKVASPTQFTLDKGDRIFYDQILTIENQWLSYKSFNG 257
 +Q G Y F+ ++ VK + + SP + G ++YD+++T G WLSY SF+G
 Sbjct: 190 SQASFDPSGT YHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSG 249

Query: 258 VRRFVLLGKASSVEKTEDKEKVSPQPQARITKTG 291
 RR++ + + + D K S + +T G
 Sbjct: 250 NRRYIPIKEPAQSVVQNDNTKPSIKVGDTVTFPG 283
 Identities = 23/73 (31%), Positives = 35/73 (47%)

Query: 103 ETKPMVEKTLPEQGN YVYSKETEVKNTPSKSAPVAFYAKKGDKVFDQVFNKDNVWISY 162
 + K + + + G Y ++ VK S +P Y + G V+YD+V W+SY
 Sbjct: 185 QLKHISQASFDPSGT YHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSY 244

Query: 163 KSFCGVRRYAAIE 175
 SF G RRY I+
 Sbjct: 245 LSFGNRRYIPIK 257

-2206-

SEQ ID 8938 (GBS91) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 7; MW 63kDa).

The GBS91-His fusion product was purified (Figure 195, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 283), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1958

A DNA sequence (GBSx2067) was identified in *S.agalactiae* <SEQ ID 6069> which encodes the amino acid sequence <SEQ ID 6070>. This protein is predicted to be thiamine biosynthesis protein. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0984(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB49673 GB:AJ248285 PROBABLE 2-DEHYDROPANTOATE 2-REDUCTASE (EC 1.1.1.169) [Pyrococcus abyssi]
Identities = 85/301 (28%), Positives = 150/301 (49%), Gaps = 7/301 (2%)

Query: 1 MLVYIAGSGAMGCRFGYQISKTNHVDVILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
M +YI G+GA+G FG ++ DV+L+ H+ AI E GLK+ G + VK+
Sbjct: 1 MKIYILGAGAIGSLFGGLLANAGEDVLLIGR-DPHVSAINKGLKIVGIKDLNVKVEATT 59

Query: 61 PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKKTIVLCLLNGLGHEDVIRQYIPEHNIL 120
E+ DLI+L TK+ L+ + I+ K + VL + NG+G+ED I ++ +
Sbjct: 60 RVPE-EKPDILIVLATKSYSTIEALKSARHIV-KGSWVLSIQNGIGNEDKIIEF--GGKAI 115

Query: 121 MGVTVWTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
G+T A ++ PG G G + ++ +V ++ N A ++ EN++
Sbjct: 116 GGITTNGAMVEAPGVIKWTGKGVTTIIGLYPQGKEKFIKVDVFNNSADIETHVSENIISW 175

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASDGVNMVHEIIEHFVTVGKAEGVELDEEEITK 240
IW KA VN +N LL+ + ++ ++M E++ E V G+E D +
Sbjct: 176 IWAKAIVNSAINPIGTLLEVKNKVIRENDFLLSMAMEVVKEGCRVALQNGIEFDVPPMDL 235

Query: 241 YVMDTSVKAHHYPSMHQDLVQNQRLEIDFLNGAVNKKGENLGIDTPYCRLITQLIHTKE 301
+ T + +Y SM QD+ + ++ TE+D++NG + + + + ++ P L+ LI KE
Sbjct: 236 F-FQTLEQTREYNYSMLQDIWRGKK-TEVDYINGKIVEYAKAVNLEAPMNLWGLIKGKE 294

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6071> which encodes the amino acid sequence <SEQ ID 6072>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1392(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2207-

Identities = 262/307 (85%), Positives = 288/307 (93%)

Query: 1 MLVYIAGSGAMGCRFGYQISKTN+DVILLDNW DHI AIKENGL VTGD E+ VKLPIMK 60
 5 Sbjct: 1 MLVYIAGSGAMGCRFGYQISKTNNDVILLDNWEDHINA IKENGLVVTGDVEETVKLPIMK 60

Query: 61 PTDATEADLIILFTKAMQLPNMLQDIKKIIGKTKVLCLLNLGHEDVIRQYIPEHNIL 120
 10 Sbjct: 61 PTEATQEADLIILFTKAMQLPQMLQDIKKIIGKETKVLCLLNLGHEDVIRQYIPEHNIL 120

Query: 121 MGVTVWTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
 Sbjct: 121 MGVTVWTAGLEGPGRAHLQGVGALNLQSM DPSNQEAGHQVADLLNEANLNATYDENVVPN 180

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVMNVHEIIEHFVTVGKAEGVELDEEEITK 240
 15 Sbjct: 181 IWRKACVNGTMNSTCALLDCTIG+LFASEDG+ MV EIIHEFV VG+AEGVEL+EEEIT+ 240

Query: 241 YVMDTSVKAHHYPSMHQDLVQNRLTEIDFLNGAVNKKGENLGIDTPYCRLITQLIHTK 300
 20 Sbjct: 241 YVMDTSVKAHHYPSMHQDLVQNRLTEIDF+NGAVN KGE LGI+TPYCR+IT+L+H K 300

Query: 301 ENVLSIK 307
 25 Sbjct: 301 EAVLNIQ 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1959

30 A DNA sequence (GBSx2068) was identified in *S.agalactiae* <SEQ ID 6073> which encodes the amino acid sequence <SEQ ID 6074>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

| | | | |
|----------|--------------------|---------------|--------------------|
| INTEGRAL | Likelihood = -3.03 | Transmembrane | 61 - 77 (61 - 78) |
| INTEGRAL | Likelihood = -1.33 | Transmembrane | 80 - 96 (79 - 96) |

35

----- Final Results -----

| | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.2211(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

40

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45

Example 1960

A DNA sequence (GBSx2069) was identified in *S.agalactiae* <SEQ ID 6075> which encodes the amino acid sequence <SEQ ID 6076>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -9.82 | Transmembrane | 317 - 333 (304 - 335) |
| INTEGRAL | Likelihood = -7.64 | Transmembrane | 187 - 203 (183 - 217) |
| INTEGRAL | Likelihood = -5.26 | Transmembrane | 24 - 40 (18 - 44) |
| INTEGRAL | Likelihood = -5.04 | Transmembrane | 143 - 159 (139 - 161) |

50

55

-2208-

```

INTEGRAL    Likelihood = -2.34    Transmembrane  116 - 132 ( 115 - 136)
INTEGRAL    Likelihood = -2.13    Transmembrane   55 -  71 (  55 -  71)
INTEGRAL    Likelihood = -0.96    Transmembrane  268 - 284 ( 268 - 284)

```

5 ----- Final Results -----

```

bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

10 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
pallidum]
Identities = 138/358 (38%), Positives = 220/358 (60%), Gaps = 18/358 (5%)

15 Query:  2  TNTVTTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQ 60
      T +++P++  F+ K+L G++ IV+ L+P AI  +  L A  H+V  Q
Sbjct:  3  TQSLSPRQ----FMMKILNGSSAGIVIGLVPPAIAAGELFRALAPLSPLFAALYHVVLPIQ 58

20 Query: 61  FFTPIMAGFLIGQQFKFTPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLI 120
      F  P + G L+G QF  +  +  +  + I SG  +  G++ + GIGD+I
Sbjct: 59  FSVPALIGTLVGLQFHCSAPEVATLAFVSVIASG-----NVTLQNGAWLITGIGDVI 110

Query: 121  NMMLTAALAVLAVKWFGNKFGLSLTIIILLPIIIGTGVGYLGWKLPLVYSYVTTLIGQGINS 180
      N+ML +ALA++ V+  K GSLTII LP+I+  G +G  LPYV +T +G+ I +
25 Sbjct: 111 NVMLISALAIILVRALRGKLSLTIIALPVIIVAVVAGGVGSFSLPYVKMITLFFVGRVIAT 170

Query: 181  FTTLQPIAMSIILAMAFSMLIVSPISTVAIGLAIGLNGMSASAASMGVASTTAVLVWATM 240
      F  LQP+ MSIL++M+FS++I+SP+S+VA+G+A+GL G+++ AA++GV+S  L+  TM
Sbjct: 171  FIALQPLILMSILLSMSFSLIIISFVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTM 230

30 Query: 241  KANKSGVPIAIALGAMKMMMPNFKHPVMAIPMLMTATVSSLTIVPLFKLVGTPASSGFGL 300
      + NK GVP+A+  GAMKM+MPN++++P++ IP+L+  V  +  LF L GTPAS+GFG
Sbjct: 231  RVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFLNLQGTASAGFGF 290

35 Query: 301  VGAVGPISAFE--AGASML---IVILSWLVIPFAVGVFVSHKICKDILKLYKDDIFVFE 353
      +G VGPI ++  A  M+  I+ L + V+ F  ++  I  D LKLY+ ++F+ E
Sbjct: 291  IGLVGPINAYRLMAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRCLKLYRRELFIPE 348

```

There is also homology to SEQ ID 1280.

40 A related GBS gene <SEQ ID 8939> and protein <SEQ ID 8940> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 8
McG: Discrim Score:      -7.24
GvH: Signal Score (-7.5): -2.94
45 Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program count: 7 value: -9.82 threshold: 0.0
INTEGRAL    Likelihood = -9.82    Transmembrane  317 - 333 ( 304 - 335)
INTEGRAL    Likelihood = -7.64    Transmembrane  187 - 203 ( 183 - 217)
50 INTEGRAL    Likelihood = -6.37    Transmembrane  143 - 159 ( 136 - 161)
INTEGRAL    Likelihood = -5.26    Transmembrane   24 -  40 (  18 -  44)
INTEGRAL    Likelihood = -2.34    Transmembrane  116 - 132 ( 115 - 136)
INTEGRAL    Likelihood = -2.13    Transmembrane   55 -  71 (  55 -  71)
INTEGRAL    Likelihood = -0.96    Transmembrane  268 - 284 ( 268 - 284)
55 PERIPHERAL Likelihood = 0.69      205
modified ALOM score: 2.46

```

*** Reasoning Step: 3

60 ----- Final Results -----

```

bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


The protein has homology with the following sequences in the databases:

ORF02147 (337 - 1359 of 1668)

EGAD|138195|TP0038(10 - 348 of 350) regulatory protein {Treponema pallidum} OMNI|TP0038 regulatory protein (pfoS/R) GP|3322295|gb|AAC65034.1|AE001189 regulatory protein (pfoS/R) {Treponema pallidum} PIR|E71373|E71373 probable regulatory protein (pfoS/R) - syphilis spirochete

%Match = 21.6

%Identity = 40.1 %Similarity = 65.6

Matches = 135 Mismatches = 112 Conservative Sub.s = 86

87 117 147 177 207 237 267 297
 LOODMGKHOSI*TKLSIIIFILIEITV*SIOHH**NNYK*N**VYKGLYILLKK*OSFLFIL*YN*LCRYE*Y*INEARY

327 357 387 417 444 474 504 534
FMTNTVTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQFFTPIMAGFLIGQQFKFT
| : | : | :: : || : | : || : : | | : : | | | : | : | : | : | :
MHTQSLSPRQFMMKILNGSAGIVIGLVPPAIAAGELFRALAPLSPFLFAALYHVVLPIQFSVPALIGTLVGLQFHCS
10 20 30 40 50 60 70

564 594 624 654 684 714 744 774
PMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLINMMLTAALAVLAVKWFGNKFGSLXIIILLPIIIGTGVG
: : : | || : :: : |||::||::| :||:: : : : ||| || ||::|
APEVATLAFVSVIASG-----NVTLLONGAWLITGIDGVINVMLISALAIILVRALRGLKLSLTIALPVIIVAVVAGG

90 100 110 120 130 140

[illegible]

1044 1074 1104 1134 1164 1194 1224 1248
 TMKANKSGVPVIAIALGAMKMMMPNFKHPVMAIPMLMTATVSSSLTVPLFKLVGTTPASSGFGLVGAVGPIASFE--AGASM
 ||: || |||: |: ||||: |||::: |: |: | : || ||||: ||::| |||| : : |
 TMRVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPTLLNGLVCGVLAWLFNMQGTPASAGFGFGLVGPINAYRLMAYTPM
 240 250 260 270 280 290 300

1269 1299 1329 1359 1389 1419 1449 1479
L---IVILSWLVIPFAVGVFVSHKICKDILKLYKDDIFVFEGQN*FGGCMLVYIAGSGAMGCRFGYQISKTNHVDVILLDNW
: |:|:|:|:| ::| |||||:|:|
VRAGILFLVYFVLSFLAAYLIDFILVDRCLKLYRRELFIPEQG
 320 330 340 350

There is also homology to SEQ ID 1276

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1961

A DNA sequence (GBSx2070) was identified in *S.agalactiae* <SEQ ID 6077> which encodes the amino acid sequence <SEQ ID 6078>. Analysis of this protein sequence reveals the following:

Possible site: 20

```
>>> Seems to have a cleavable N-term signal seq.
```

----- Final Results -----

```
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07127 GB:AP001518 thioredoxin reductase [Bacillus halodurans]

Identities = 163/325 (50%), Positives = 222/325 (68%), Gaps = 3/325 (0%)

-2210-

Query: 5 IYDITIVGGGPVGLFAAFYAGLRGVSVKIIESLSSELGGQPAILYPEKKIYDIPGYPVITG 64
 +YDITI+GGGP GLFAAFY G+R VKIIES+ +LGGQ A LYPEK IYD+ G+P +
 Sbjct: 7 LYDITIIGGGPTGLFAAFYGGMRQAKVKIIESMPQLGGQLAALYPEKYIYDVAGFPKVK 66

Query: 65 RELIDKHIEQLERFKDSIEICLKEEVLSEFEK-VDDVFTIQTDKDQHLRAIVFACGNGAF 123
 ++L++ Q E+F +I L++ V + K DD FTI+TDK+ H S+AI+ G GAF
 Sbjct: 67 QDLVNDLKRQAEQFNPTI--ALEQSVQNVTKETDDFTTIKTDKETHYSKAIITAGAGAF 124

Query: 124 APRLLGLENEENYADNNLFYNVTLEQFAGKHVVICGGGDSAVDWANELDKIAASVAIVH 183
 PR L +E + Y NL Y V L +AGK+V+I GGGDSAVDWA L+ +A +V ++H
 Sbjct: 125 QPRRLEVEGAKQYEGKNLQYFVNDLNAYAGKNVLISGGGDSAVDWALMLEPVAKNVTLIH 184

Query: 184 RRDAFRAHEHSVDILKASGVRIITPYVPIGLNGDSQVRSSLVVQKVKGEVIELPLDNL 243
 RRD FRAHEHSV++L+ S V ILTP+ L+GD +++ + +Q+VKGD V L +D +I
 Sbjct: 185 RRDKFAHEHSVELLQKSSVNILTPFAISELSGDGEKIHHTVQEVKGDVAVETLDVDEVI 244

Query: 244 VSGFGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEA 303
 V+FGF +S ++ W L+ +++SI V++ ET G+YA GD YPGKV+LIATG+GEA
 Sbjct: 245 VNFGFVSSLGPIKGWGLEIEKNSIVVNTKMETNIPGIYAAGDICTYPGKVLIATGFGGEA 304

Query: 304 PVAINQAINIYIPDRDNRVHSTSL 328
 P A+N A +I P HSTSL
 Sbjct: 305 PTAVNNAKAFIDPTARVFPGHSTSL 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6079> which encodes the amino acid sequence <SEQ ID 6080>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 (8 - 24)

----- Final Results -----
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15201 GB:Z99120 similar to thioredoxin reductase [Bacillus subtilis]
 Identities = 173/328 (52%), Positives = 223/328 (67%), Gaps = 4/328 (1%)

Query: 4 KAYDITIIGGGPIGLFAAFYAGLRGVTVKIIESLSSELGGQPAILYPEKMIYDIPAYPSLT 63
 K YDITIIGGGP+GLF AFY G+R +VKIIESL +LGGQ + LYPEK IYD+ +P +
 Sbjct: 6 KVDITIIGGGPVGLFTAFYGGMRQASVKIIESLPQLGGQLSALYPEKYIYDVAGFPKIR 65

Query: 64 GVELTENLIKQLSRFEDRTTICLKEEVLTFDKVKGKGG-FSIRTNKAHFHFSKAIITACNGA 122
 EL NL +Q+++F+ TICL++ V + +K G F + K K I GNGA
 Sbjct: 66 AQELINNLKEQMAKFDQ--TICLEQAVESVEKQADGVFKLVQMKPTTLKRSCITAGNGA 123

Query: 123 FAPRTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALALEEDIAESVTVV 182
 F PR L LE+ E + NL Y V L +FAG++V I GGGDSAVDWAL LE IA+ V+++
 Sbjct: 124 FKPRKLELENAEQYEGKNLHYFVDDLQKFAGRRAVAILGGGDSAVDWALMLEPIAKEVSII 183

Query: 183 HRRDAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAEKLVQKVEDEVLELELDSL 242
 HRRD FRAHEHSVE L AS VN+LTP+VP L G + E+LV+++VK D LE+D L
 Sbjct: 184 HRRDKFRAHEHSVENLHASKVNVLTPFVPAELIGEDKI-EQLVLEEVKGRKEILEIDDL 242

Query: 243 IVSFGFGFSTSNKNLKNWNLDYKRSSITVSPFLFQTSQEGIFAIGDAAAYNGKVDLIATGFGE 302
 IV++GF +S +KNW LD +++SI V +T+ EG FA GD Y GKV+LIA+GFGE
 Sbjct: 243 IVNYGFVSSLGPIKNWGLDIEKNSIVVKSTMETNIEGFFAAGDICTYEGKVNLIASGFGE 302

Query: 303 APTAVNQAINIYIPDRDNRVHSTSLID 330
 APTAVN A Y+ P + +HSTSL +
 Sbjct: 303 APTAVNNAKAYMDPKARVQPLHSTSLFE 330

An alignment of the GAS and GBS proteins is shown below.

-2211-

Identities = 242/324 (74%), Positives = 279/324 (85%)

Query: 6 YDITIVGGGPVGLFAAFYAGLRGVSVKIIIESLSELGGQPAILYPEKKIYDIPGYPVITGR 65
 YDITI+GGGP+GLFAAFYAGLRGV+VKIIIESLSELGGQPAILYPEK IYDIP YP +TG
 5 Sbjct: 6 YDITIIGGGPIGLFAAFYAGLRGVTVKIIIESLSELGGQPAILYPEKMIYDIPAYPSLTGV 65

Query: 66 ELIDKHIEQLERFKDSIEICLKKEEVLSEFEKVDVFTIQTQDKDQHLRAIVFACGNGAFAP 125
 EL + I+QL RF+D ICLKEEVL+F+KV F+I+T+K +H S+AI+ ACGNGAFAP
 10 Sbjct: 66 ELTENLIKQLSRFEDRTTICLKKEEVLTFDKVKGGSIRTNKAHFASKAIIACGNGAFAP 125

Query: 126 RLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWANELDKIAASVAIVHRR 185
 R LGLE+EEN+AD+NLFYNV +L+QFAG+ VVICGGGDSAVDWA L+ IA SV +VHRR
 15 Sbjct: 126 RTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALAEEDIAESVTVVHRR 185

Query: 186 DAFRAHEHSVDILKASGVRIITPYVPVIGLNGDSQVSSLVQKVGDEVIELPLDNLIVS 245
 DAFRAHEHSV++LKAS V +LTPYVP L G LV+QKVK DEV+EL LD+LIVS
 20 Sbjct: 186 DAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAELKVIQKVEDEVLELELDSLIVS 245

Query: 246 FGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEAPV 305
 FGFSTSNKNL+ WNLDYKRSSI VS LF+T+QEG++AIGDAA Y GKV+LIATG+GEAP
 25 Sbjct: 246 FGFSTSNKNLKNWNLDYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGGEAPT 305

Query: 306 AINQAINIYIPDRDNRVHSTSLI 329
 A+NQAINIYIPDRDNRVHSTSLI
 25 Sbjct: 306 AVNQAINIYIPDRDNRVHSTSLI 329

SEQ ID 6078 (GBS178) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 5; MW 37.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 8; MW 62.4kDa).

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1962

A DNA sequence (GBSx2071) was identified in *S.agalactiae* <SEQ ID 6081> which encodes the amino acid sequence <SEQ ID 6082>. This protein is predicted to be tRNA methyltransferase (trmD). Analysis of
 35 this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1496(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BAB06198 GB:AP001515 tRNA methyltransferase [Bacillus halodurans]
 Identities = 144/246 (58%), Positives = 186/246 (75%), Gaps = 6/246 (2%)

Query: 2 MKIDILTLFPPEMFAPLEHS-IVGKAKERGLLEINYHNFRENAE-KSRHVDDEPYGGGQGM 59
 MKID LTLFPPEMF + HS I+ +A+ERG + NFRE +E K + VDD PYGGG GM
 50 Sbjct: 1 MKIDFLTLPPEMFQGVILHSSILKQAQERGAVSFRVNVFREYSENKHKKVDDYPYGGGAGM 60

Query: 60 LLRAQPIFDITDKIDAQKA---RVILLDPAGRITFDQDFAEELSKEDELIFICGHYEGYDE 116
 +L QP+FD ++ + + RVIL+ P G TF Q AEEL++ + LI +CGHYEGYDE
 55 Sbjct: 61 VLSFPQLPDAVEDLTKKSSSTPRVILMCPQGETFTQKAEELAQAELHLLCGHYEGYDE 120

Query: 117 RIKS-LVTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDSSFSSGLLEYPQ 175
 RI+S LVTDE+S+GD+VLTGGEL AM + D+ RL+P V+G ETS Q DSFS+GLLEYPQ
 Sbjct: 121 RIRSYLVTEDELSIGDYVLTGGELGAMVIADSVTRLLPAVLGNETSQAQTSFSTGLLEYPQ 180

-2212-

Query: 176 YTRPYDYLGMTPVDVLMMSGHHENIRKWRLEQSLRKTLERRPDLLLENYAMTDEERLILEKI 235
 YTRP D+ G VPDVL+SGHH+NI +WR EQSL++TLERRPDLL +T+EE+ +L+ I
 Sbjct: 181 .YTRPADFRGWKVPDVLVSGHHQNIERWRKEQSLKRTLERRPDLLLEGRKLTETEEQELLDLSI 240

5 Query: 236 KTEIER 241
 + + E+
 Sbjct: 241 RKQQEK 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6083> which encodes the amino acid
 10 sequence <SEQ ID 6084>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2705(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 195/240 (81%), Positives = 224/240 (93%)

Query: 2 MKIDILTLFPEMFAPLEHSIVGKAKERGLLEINYHNFRENAESRSHVDDEPYGGGQGMML 61
 MKIDILTLFPEMFAPLEHSIVGKAKE+GLL+I+YHNFR+ AEK+RHVDDEPYGGGQGMML
 Sbjct: 1 MKIDILTLFPEMFAPLEHSIVGKAKEKGLLDIHYHNFRDYAEKARHVDDEPYGGGQGMML 60

25 Query: 62 RAQPIFDITDKIDAQKARVILLDPAGRTFDQFAEELSKEDLIFICGHYEGYDERIKSL 121
 RAQPIFDTI++I+A+K R+ILLDPAG+ F Q +AEEL+ E+ELIFICGHYEGYDERIK+L
 Sbjct: 61 RAQPIFDTIEQIEAKKPRIILLDPAGKPFTQAYAEELALEEELIFICGHYEGYDERIKTL 120

30 Query: 122 VTDEVSIGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQYTRPYD 181
 VTDE+SLGDFVLTGGELAAMTMVDATVRLIP+V+GKE+SHQDDSFSSGLLEYPQYTRPYD
 Sbjct: 121 VTDEISLGDFVLTGGELAAMTMVDATVRLIPQVLGKESSHQDDSFSSGLLEYPQYTRPYD 180

35 Query: 182 YLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLLENYAMTDEERLILEKIKTEIER 241
 Y GMTVPDVLMSGHHE IR WRLE+SL+KT RRPDLLE+Y ++EER +L+KIK +++
 Sbjct: 181 YRGMTVPDVLMSGHHERIRLWRLEESLKKTYLRRPDLLLEHYNFSEERKLLDKIKEALDQ 240

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

40 Example 1963

A DNA sequence (GBSx2072) was identified in *S.agalactiae* <SEQ ID 6085> which encodes the amino
 acid sequence <SEQ ID 6086>. This protein is predicted to be 16S rRNA processing protein. Analysis of
 this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.71 Transmembrane 32 - 48 (32 - 52)

----- Final Results -----

50 bacterial membrane --- Certainty=0.2084(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9381> which encodes amino acid sequence <SEQ ID 9382>
 was also identified.

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13475 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]

-2213-

Identities = 88/174 (50%), Positives = 128/174 (72%), Gaps = 1/174 (0%)

Query: 54 VTMEYFNVGKIVNTQGLQGEMRVLSVTDVFVEERFKKGQVLALFDEKNQFVMDIEIASHRK 113
 +T +FNVGKIVNT G++GE+RV+S TDF EER+K G L LF + +++ + +HR
 Sbjet: 1 MTKRWFNVGKIVNTHGKGEVRVISKTDFAEERYKPGNTLYLFMDGRNEPVEVTVNTHRL' 60

Query: 114 QKNFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYHHEIIGLDVYEGE-ELIG 172
 K F +++FK ++N++E+ K +KV E++L +L +GEFY+HEIIG +V+ E ELIG
 Sbjet: 61 HKQFHLLQFKERQNLNEVEELKNAIKVPPEELGELNEGEFYFHEIIGCEVFTTEEGELIG 120

Query: 173 KIKEILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDE 226
 K+KEIL PGANDVWV+ R GK+D L+PYI VV +D+ +++++ELMEGL DE
 Sbjet: 121 KVKEILTPGANDVWVIGRKGGKDALIPYIESVVKHIDVREKKIEIELMEGLIDE 174

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6087> which encodes the amino acid sequence <SEQ ID 6088>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2787(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 133/172 (77%), Positives = 153/172 (88%)

Query: 56 MEYFNVGKIVNTQGLQGEMRVLSVTDVFVEERFKKGQVLALFDEKNQFVMDIEIASHRKQK 115
 MEYFNVGKIVNTQGLQGEMRVLSV+DF EERFKKG LALFD+K++FV ++ I SHRKQK
 Sbjet: 1 MEYFNVGKIVNTQGLQGEMRVLSVSDFAEERFKKGSQALALFDDKDRFVQEVTVSHRKQK 60

Query: 116 NFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYHHEIIGLDVYEGEELIGKIK 175
 +FDIIKFK MYHIN IEKYKG+TLKV++D DL++GEFYH+IIG+ VYE + LIG +K
 Sbjet: 61 HFDIIKFKDMYHINAIEKYGYTLKVSQDNQDLQGEFYHQTIGMAVYEKDVLIHVK 120

Query: 176 EILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDED 227
 EILQPGANDVW+V+R GKRDLPLPYIPPVVL VD+ N+RV VELMEGLDDED
 Sbjet: 121 EILQPGANDVWIVKRQGRDLPLPYIPPVVLNVDVFNKRVDELMEGLDDED 172

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1964

A DNA sequence (GBSx2073) was identified in *S.agalactiae* <SEQ ID 6089> which encodes the amino acid sequence <SEQ ID 6090>. This protein is predicted to be similar to *E. coli* ykfc (11). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9909> which encodes amino acid sequence <SEQ ID 9910> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2214-

>GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
Identities = 366/425 (86%), Positives = 396/425 (93%)

```

5 Query: 12 MSELDDKILSRNMLEAYKQVKSNGKSAGINGVTIEQMDDYLHQNWRQTKQLIKERSYKP 71
    MS+LLDKILSR NMLEAY QVKSNGKSAGI+G+TIE+MD+YL QNWR TK+LIK+R YKP
Sbjct: 1 MSKLLDKILSRNMLEAYNQVKSNGKSAGIDGMTIEEMDNYLRQNWRLTKELIKQRKYKP 60

Query: 72 QPVLRLVEIPKPNGGVRNLGIPTAMDRMIQQAIVQVLSPLCEKHFSEYSYGFRPNRSCETA 131
    QPVL+VEIPKP+GG+R LGIPT MDRMIQQAIVQV+SP+CE HFS+ SYGFRPNRSCE A
10 Sbjct: 61 QPVLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKA 120

Query: 132 IVQLLEYLNDGYEWIVDIDLEKFFDTVPQDRMLSLVHNIIQDGDTESLIRKYLHSGVVIN 191
    I++LLEYLNDGYEWIVDIDLEKFFDTVPQDRMLSLVHNII+DGDTESLIRKYLHSGV+IN
15 Sbjct: 121 IMKLEYLNDGYEWIVDIDLEKFFDTVPQDRMLSLVHNIIEDGDTESLIRKYLHSGVIIN 180

Query: 192 GQRHKTIVGTTPQGGNLSPLLSNIMLNELDKGLEKRGLRFVRYADDCVITVGSEAAAKRVM 251
    GQR+KTIVGTTPQGGNLSPLLSNIMLNELDK LEKRGLRFVRYADDCVITVGSEAAAKRVM
Sbjct: 181 GQRYKTIVGTTPQGGNLSPLLSNIMLNELDKLEKRGLRFVRYADDCVITVGSEAAAKRVM 240

20 Query: 252 HSVSSYIEKRLGLKVNMTTKTKIVRPNKLYLGFGEKSPKWKCRPHQDSVQSFKRKLKQ 311
    +SVS +IEKRLGLKVNMTTKTKI RP +LKYLGFGFWKS GWK RPHQDSV+ FK KLK+
Sbjct: 241 YSVSRFIEKRLGLKVNMTTKTKITRPRELKYLGFGFWKSSDGWKS RPHQDSVRRFKLKLK 300

Query: 312 LTMRKWSIDLITRIERLNWVIRGWINYFSLGNMKSIMTQIDERLRTIRVVIWKQWKKKA 371
    LT RKWSIDL RIE+LN IRGWINYFSLGNMKSII+ IDERLRTIR+R+IIWKQWKKK+
25 Sbjct: 301 LTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRTLRMIWKQWKKKS 360

Query: 372 KRLWGLLKLGVARWIADKVSQWGDHYQLVAQKSVLKRAISKPALAKRGLVSCLDYYLERH 431
    +RLWGLLKLGV +WIADKVSQWGDHYQLVAQKSVLKRAISK L KRGLVSCLDYYLERH
30 Sbjct: 361 RRLWGLLKLGVPKWIADKVSQWGDHYQLVAQKSVLKRAISKPVLEKRGLVSCLDYYLERH 420

Query: 432 ALKVS 436
    ALKVS
Sbjct: 421 ALKVS 425
35

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1965

A DNA sequence (GBSx2074) was identified in *S.agalactiae* <SEQ ID 6091> which encodes the amino acid sequence <SEQ ID 6092>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -0.37    Transmembrane    7 - 23 ( 7 - 23)
45 ----- Final Results -----
                bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 821> which encodes the amino acid sequence <SEQ ID 822>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
55    INTEGRAL    Likelihood = -2.87    Transmembrane    1157 -1173 (1157 -1174)

----- Final Results -----
                bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2215-

An alignment of the GAS and GBS proteins is shown below.

Identities = 1031/1064 (96%), Positives = 1042/1064 (97%)

| | | |
|----|------------|---|
| 5 | Query: 1 | MRKKQKLPPFDKLAIALISTSILLNAQSDIKANTVTEDTPATEQAVEPPQPIAVSEESPSS 60 |
| | Sbjct: 1 | LRKKQKLPPFDKLAIALMSTSIILLNAQSDIKANTVTEDTPATEQAVETPQPTAVSEEAPSS 60 |
| 10 | Query: 61 | KETKTSQTPSDVGETVADDANDLAPQAPAKTADTPATSKATIRDLNDPSHVKTLOEKAGK 120 |
| | Sbjct: 61 | KETKT QTP D ET+ADDANDLAPQAPAKTADTPATSKATIRDLNDPS VKTLQEKAGK 120 |
| 15 | Query: 121 | GVTGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKENLEKAKKEHGITYGEWVNDKVAYYHD 180 |
| | Sbjct: 121 | G GTVVAVIDAGFDKNHEAWRLTDKTKARYQSKE+LEKAKKEHGITYGEWVNDKVAYYHD 180 |
| 20 | Query: 181 | YSKDGKNAVDQEHGTHVSGILSGNAPSEMKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240 |
| | Sbjct: 181 | YSKDGK AVDQEHGTHVSGILSGNAPSE KEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240 |
| 25 | Query: 241 | NYAQAIRDAVNLAGAKVINMSFGNAALAYANLPDETKKAFDYAKSGVSIIVTSAGNDSSFG 300 |
| | Sbjct: 241 | NYAQAI DAVNLAGAKVINMSFGNAALAYANLPDETKKAFDYAKSGVSIIVTSAGNDSSFG 300 |
| 30 | Query: 361 | GKPRPLADHPDYGVVGTCAAADSTLTVAASYSPDKQLTETATVKTDDHQDKEMPVLSTNR 360 |
| | Sbjct: 361 | GK RPLADHPDYGVVGTCAAADSTLTVAASYSPDKQLTETATVKT D QDKEMPVLSTNR 360 |
| 35 | Query: 421 | KGFPIELPNVDQMPAAAFISRRDGLLLKDNPKQKTTTFNATPKVLPTASGKLSRFSSWGLT 480 |
| | Sbjct: 421 | KGFPIELPNVDQMPAAAFISR+DGLLLK+NPQKTTTFNATPKVLPTASGKLSRFSSWGLT 480 |
| 40 | Query: 481 | ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPMTPSE 540 |
| | Sbjct: 481 | ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPMTPSE 540 |
| 45 | Query: 541 | RDLAKKVLMSATALLYDEDEKAYFSPRQQGAGAVDAKASAAATMYVTDKNTSSKVHLN 600 |
| | Sbjct: 541 | RDLAKKVLMSATALLYDEDEKAYFSPRQQGAGAVDAKASAAATMYVTDKNTSSKVHLN 600 |
| 50 | Query: 601 | NVSDKFEVTVTVHNKSDKPQELYYQTVQTDKVDGKHFALAPKALYETSWQKITIPANSS 660 |
| | Sbjct: 601 | NVSDKFEVTVTVHNKSDKPQELYYQ TVQTDKVDGK FALAPKALYETSWQKITIPANSS 660 |
| 55 | Query: 661 | KQVTVPIDASRFSKDLLAMKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720 |
| | Sbjct: 661 | KQVT+PID S+FSKDLLA MKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720 |
| 60 | Query: 721 | KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN 780 |
| | Sbjct: 721 | KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN 780 |
| 65 | Query: 781 | IEDIESSEITETIFAGTFAKQDDDSHYIHRHANGKPYAAISPNGDGNRDYVQFGTFLR 840 |
| | Sbjct: 781 | IEDIESSEITETIFAGTFAKQDDDSHYIHRHANGKPYAAISPNGDGNRDYVQFGTFLR 840 |
| | Query: 841 | NAKNLVAEVLKKEGNVWVTSEVTEQVVKKNYNNDLASTLGSTRFEKTRWDGKNDKGKVVAN 900 |
| | Sbjct: 841 | NAKNLVAEVLKKEGNVWVTSEVTEQVVKKNYNNDLASTLGSTRFEKTRWDGK+KDGKVVAN 900 |
| | Query: 901 | GTYYTYRVRYTPISSGAKEQHTDFDVIDNTTPEVATSATFSTEDSRLTLASKPKTSQPVY 960 |
| | Sbjct: 901 | GTYYTYRVRYTPISSGAKEQHTDFDVIDNTTPEVATSATFSTEDRRLTLASKPKTSQPVY 960 |

-2216-

Query: 961 RERIAITYMDEDLPTTEYISPNEGDTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020
 RERIAITYMDEDLPTTEYISPNEGDTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT
 5 Sbjct: 961 RERIAITYMDEDLPTTEYISPNEGDTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020

Query: 1021 YTPVTKLLEGHHSNKPEQDGSQAPDKKPEAKPEQDGSQGPDKK 1064
 YTPVTKLLEGHHSNKPEQDGSQAPDKKPE KPEQDGSQ GPDKK
 Sbjct: 1021 YTPVTKLLEGHHSNKPEQDGSQAPDKKPEKPEQDGSQGPDKK 1064

10 A related GBS gene <SEQ ID 8941> and protein <SEQ ID 8942> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10
 McG: Discrim Score: 5.69
 15 GvH: Signal Score (-7.5): -3.33
 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -0.37 threshold: 0.0
 INTEGRAL Likelihood = -0.37 Transmembrane 7 - 23 (7 - 23)
 PERIPHERAL Likelihood = 2.81 508
 20 modified ALOM score: 0.57

*** Reasoning Step: 3

----- Final Results -----
 25 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8942 (GBS276) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 30 extract is shown in Figure 46 (lane 2; MW 123kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 5; MW 46.5kDa).

The GBS276-His fusion product was purified (Figure 206, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 296), which confirmed that the protein is immunoaccessible on GBS bacteria.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1966

A DNA sequence (GBSx2075) was identified in *S.agalactiae* <SEQ ID 6093> which encodes the amino acid sequence <SEQ ID 6094>. Analysis of this protein sequence reveals the following:

40 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.4286(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2217-

Example 1967

A DNA sequence (GBSx2076) was identified in *S.agalactiae* <SEQ ID 6095> which encodes the amino acid sequence <SEQ ID 6096>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 30
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood =-11.15    Transmembrane    19 - 35 ( 11 - 39)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9911> which encodes amino acid sequence <SEQ ID 9912> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6096 (GBS654) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 8 & 10; MW 51.2kDa + lane 9; MW 27kDa). Purified GBS654-GST is shown in Figure 245, lane 11.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1968

A DNA sequence (GBSx2077) was identified in *S.agalactiae* <SEQ ID 6097> which encodes the amino acid sequence <SEQ ID 6098>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 14
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
30  bacterial cytoplasm --- Certainty=0.4174 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9913> which encodes amino acid sequence <SEQ ID 9914> was also identified.

35 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF27324 GB:AF178424 unknown [Lactococcus lactis]
  Identities = 26/75 (34%), Positives = 45/75 (59%), Gaps = 4/75 (5%)

40  Query: 11 MAFEPKNSLTKVLKES-IDEEKKEIFSSSEMNI RDFERTKQYQFTLQPSVRKKIDRLSKE 69
      MAF+  + ++ VL S L + K E+      I E K Y FTL+PSV++ +++L+++
  Sbjct: 1  MAFDVDDKKVKTVLSNSSLAKSKVEL---PKKIESEENKKSYSFTLEPSVKEGLEKLAEK 57

  Query: 70 KGYRSASSFINDFPK 84
      + Y++ S F+ND K
45  Sbjct: 58 QNYKNTSQFLNDLIK 72

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-2218-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1969

A DNA sequence (GBSx2078) was identified in *S.agalactiae* <SEQ ID 6099> which encodes the amino acid sequence <SEQ ID 6100>. This protein is predicted to be ParA. Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF27325 GB:AF178424 ParA [Lactococcus lactis]
Identities = 49/104 (47%), Positives = 72/104 (69%)

Query: 22 L SERLEEFKTEAFDFKTRASYVTAKLFFLGNMIKHNTNSSKELIRSLKNDKSVLAMIPHK 81
L ERL+ FK E D +TR +Y+TA +F+GN I+HNT SS+E + DK +AMIP K
Sbjct: 157 L IERLQNFKDEVIDARTRETYITAIPYFVGNRIRHNTKSSREFSEKISQDKGTIAMIEPK 216

Query: 82 ELFNIRSTLDKKSLSYMSDKELYSRDSKFFKEIDFTFRKITDKL 125

ELFNIRSTLD L M DK++++ + F++++F F +IT+K+
Sbjct: 217 ELFNIRSTLDGVPLVEMEKDKDVFNSNKVFYEKLNFAFNEITNKI 260

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1970

A DNA sequence (GBSx2079) was identified in *S.agalactiae* <SEQ ID 6101> which encodes the amino acid sequence <SEQ ID 6102>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2830 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1971

A DNA sequence (GBSx2080) was identified in *S.agalactiae* <SEQ ID 6103> which encodes the amino acid sequence <SEQ ID 6104>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

-2219-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2618(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 242/259 (93%), Positives = 249/259 (95%)

10 Query: 1 MCRWLNMPHSSYYYQAVESVSETEFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
 MCRWLN+P SSSYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
 Sbjct: 1 MCRWLNIPRSSYYYKAVEPVSEAELEESIKAIFLESKARYGSRKIKICLNNEGITLSRRR 60

15 Query: 61 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDROQFKQERPLQALVTDLTIVRVGNR 120
 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDROFK ERPLQALVTDLTIVRVGNR
 Sbjct: 61 IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDROFKPERPLQALVTDLTIVRVGNR 120

20 Query: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYALTQVVMFHSRDXKEFDNQLID 180
 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPY LTKVVMFHSR KEF+NQLID
 Sbjct: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYPLTKVVMFHSRDXKEFNNQLID 180

25 Query: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFOLEELALKTKDYVHWWNY 240
 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFO LEELALKTK YVHWWNY
 Sbjct: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFOLEELALKTKAYVHWWNY 240

30 Query: 241 HRIHGSLNYQTPMTKRLIA 259
 HRIHGSLNYQTPMTKRLIA
 Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259

There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1972

35 A DNA sequence (GBSx2081) was identified in *S.agalactiae* <SEQ ID 6105> which encodes the amino acid sequence <SEQ ID 6106>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3325(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1973

50 A DNA sequence (GBSx2082) was identified in *S.agalactiae* <SEQ ID 6107> which encodes the amino acid sequence <SEQ ID 6108>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

-2220-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9917> which encodes amino acid sequence <SEQ ID 9918> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD44095 GB:AF115103 orf359 gp [Streptococcus thermophilus
bacteriophage Sfi21]
Identities = 92/357 (25%), Positives = 162/357 (44%), Gaps = 33/357 (9%)

Query: 45 RKNQYGKTFETMKEAYDELVRKYEFANKVSLNMTFENYMNKIYLRAYKQK-VQSVT 103
RK + F T EA ++ + V+++ ++T +Y K + YK+ V +T
Sbjct: 24 RKPKTGGGFRFKSEAIKAAEMELKLQDNVNVDE-DITLYDYF-KQWCEVYKKPTVSKIT 81

Query: 104 YKTALPHHKLFIQYFGLKPLKAITPRDCEAFRLHIIENYSENYAKNLWSRF-----KACMG 159
YK + + +FG K LK+IT + + ++ +Y++ +A++ RF KAC+
Sbjct: 82 YKAYINSQRKIELFFGDKKLKSITATEYQ----RVLNSYAKTHAQDTVERFNVHVKACIE 137

Query: 160 YAERLGYISNMPCKALD---NPRGKHPETPFWTYAEFQTFIKSFDLHDYEELQRFTAIWL 216
A GYI CK +G+ ET F E++ I ++ + E + A+++
Sbjct: 138 MAVHEGYIKRNFCKFAKINAKNKGRIETKFLEVEEYERLI--YETSKHPEYASYAALYI 195

Query: 217 YYMTGVRVSEGLSLCWEDIDFDKKFLKVHTTLEKDENGNNWYRKDQTKTPAGERLIELDDI 276
TG+R +E L L +DI D L V+ T + N + TKT + R I LDD
Sbjct: 196 IAKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDYKNNTGFM---PTKTKSSIREIPLDDE 252

Query: 277 TIEVLQVWRKNQFANQDQDFIISRFQDPFCKSTICRIIKRKAQQVGVVITGKGLRHS 336
I + +Q D I+ + T+ +I+ R+ + LRH++A
Sbjct: 253 FINFI-----DQLPPTDDGRILPSLSNNAVNKTLRKIVGRE-----VRVHSLRHTYA 299

Query: 337 SYLINVLKDKILYVARRMGHADKSTTLNTYSHWFNLDKTVSEETQNIKSAGLDSI 393
SYLI D++ V++ +GH + + TL Y+H E+I Q G +++
Sbjct: 300 SYLI-AHDIDLISVSQVLGHENLNITLEVYAHQLQEQKSRNDEKIQMWTECGRNAL 355

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6109> which encodes the amino acid sequence <SEQ ID 6110>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5549(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/127 (87%), Positives = 119/127 (93%)

Query: 242 LKVHTTLEKDENGNNWYRKDQTKTPAGERLIELDDITIEVLQVWRKNQFANQDQDFIISRF 301
LKVHTTLEKDENGNNWYRKDQTKTPAGERLIELDD+TI VL+ WR+NQ N DTDFIISRF
Sbjct: 1 LKVHTTLEKDENGNNWYRKDQTKTPAGERLIELDDVTIVLENWRRNQVNTDTDFIISRF 60

Query: 302 GDFPFCKSTICRIIKRKAQQVGVVITGKGLRHSASYLINVLKDKILYVARRMGHADKST 361
G+PFCKSTICR+IK KAQ +GVPVITGKGLRHS+ASYLINVLKDKILYVA+ MGHADKST
Sbjct: 61 GEPFCKSTICRVIKHKAQSIGVPVITGKGLRHSASYLINVLKDKILYVAKCMGHADKST 120

Query: 362 TLNTYSH 368
TLNTYSH
Sbjct: 121 TLNTYSH 127

-2221-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1974

A DNA sequence (GBSx2083) was identified in *S.agalactiae* <SEQ ID 6111> which encodes the amino acid sequence <SEQ ID 6112>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1975

A DNA sequence (GBSx2084) was identified in *S.agalactiae* <SEQ ID 6113> which encodes the amino acid sequence <SEQ ID 6114>. This protein is predicted to be repressor protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9919> which encodes amino acid sequence <SEQ ID 9920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98432 GB:L29324 repressor protein [Streptococcus pneumoniae]
Identities = 38/65 (58%), Positives = 52/65 (79%), Gaps = 1/65 (1%)

35 Query: 2 MYRRLRDRLREDNDFQKYVAEK-LSFTHSAYSKIERGERILSADVILKLSNLYNVSTDYL 60
M +R+RDLRED+D TQ+YVA+ L+ T SAYSK+E G R++S D +IKL++ YNVS DYL
Sbjct: 1 MLKRIRDLREDDDLTQEYVAKTILNCTRSAYSKMESGTRILISIDDLIKLADFYNVSLDYL 60

40 Query: 61 LGQTD 65
+G+ D
Sbjct: 61 VGRVD 65

There is also homology to SEQ ID 582.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2222-

Example 1976

A DNA sequence (GBSx2085) was identified in *S.agalactiae* <SEQ ID 6115> which encodes the amino acid sequence <SEQ ID 6116>. This protein is predicted to be relaxase. Analysis of this protein sequence reveals the following:

5 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC98434 GB:L29324 relaxase [Streptococcus pneumoniae]
 Identities = 223/417 (53%), Positives = 310/417 (73%), Gaps = 5/417 (1%)

 Query: 1 MVITKHYAVHGKKYRRQLIKYIILDPKTRNLSLISDFGMSNYLDFPDYVELVKMYQNNFL 60
 MVITKH+A+HGK YR +LIKYL+P KT+NL+L+SDFGM NYLDFP Y ELVKMY +NFL
 Sbjct: 1 MVITKHFAIHGKNYRSKLIKYLNPSTKNTLTVSDFGMRNYLDFPSPYKELVKMYNDNFL 60

20 Query: 61 SNDQLYDSRFRDQEKQKQKIHAAHHIIQSFSPEDKLSPEEINRIGYETIKELIGGQYKFIV 120
 SND LY+ R DRQE Q+KIH+HHIIQSFSFSP+D L+PE+INRIGYE KEL GG+++FIV
 Sbjct: 61 SNDTLYEYFRHQRQEVNQKRIHSHHIIQSFSPPDHLTPQINRIGYEAKELTGGRFRFIV 120

25 Query: 121 ATHVDQDHCHNHIIINSINSQSQKKLKWDYALERNLQMSDRISKVAGAKIIPPKRYSHR 180
 ATHVD+ H HNHII+NSI+ S KK WDY E NL+M+SDR+SK+AGAKII RYSHR
 Sbjct: 121 ATHVDKGHIHNNHIIILNSIDQNSDKKFLWDYKAEHNLRMVSDRLSKIAGAKII-ENRYSHR 179

30 Query: 181 DYEYVRRSNHKYELKQRLFFLMEHSIDFNDFMQAEQLNVKIDFSRKHSRFFMTDRNMKQ 240
 YEYVR++N+KYE+KQR++FL+E+S +F D +KA+ L++KIDF KH +FMTD NMKQ
 Sbjct: 180 QYEVYRKTNYKYEIKQRVYFLIENSKNFEDLKKKAKALHLKIDFRHKHVTYFMTDSNMKQ 239

35 Query: 241 VIQGDKLNKREPYSKEYFQRYFAKKKIELILEFLLLRNSFDDLVEKARLLGLELKS KKK 300
 V++ KL++++PY++ YF++ F +++I ILEFLL + + ++L+++A + GL++ K+K
 Sbjct: 240 VVRDSKLSRKQPYNETYFEKKFVQREIINILEFLLPKMKNMNELIQRAEVFGLKIIPKEK 299

40 Query: 301 TIDFVLSDGKSCISIPNKSRLKKNLYDTTYFDSYFKEHDVFVFLHNEVKIEFEKFETQQ 360
 + F DG I + + L K NLY +YF YF + VL N + + + + +
 Sbjct: 300 HVLFEF-DG---IKLAEQELVKS NLYSVSYFQDYFNNKNETFVLNKNLVELYNEEKI K 355

 Query: 361 LSEILTVEEITEAYETYKTRDAVHEFEVEITEEQIEKIVLDGLFVKVWVGIGQEG 417
 E+ + E + ++Y+ +K RDAVHEFEVE+ QIE++V G+++KV GI ++ L
 Sbjct: 356 EKELPSEEMVWKS YQDFKRNRDAVHEFEVELNLNQIEEVVEHGIIYIKVQFGIDKKDL 412

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6117> which encodes the amino acid sequence <SEQ ID 6118>. Analysis of this protein sequence reveals the following:

 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3114(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

 Identities = 103/218 (47%), Positives = 170/218 (77%)

 Query: 393 EEQIEKIVLDGLFVKVWVGIGQEG LIFIPNHQNLNILEQENKKQYQVFIRETSSYFIYHKE 452
 E QIE+++ + +++KV + Q GLIFIPN+QL+I ++EN K+Y+V+IRET+ +FIY+KE
 Sbjct: 2 EHQIERLIAEDIYIKVSFSVKQSG LIFIPNYQLDIRKEENHKKYKVYIRETAQFFIYNKE 61

-2223-

Query: 453 DSEMNRFMKGRDLIRQLTFDNKSLPYKRRISLVSLQOKIEEINLLMTLNIONKSFLELKD 512
 SE+NR+M+G +LI QLT D+KS+P +RR ++ +L++KIEEI+LL+ L+ +NK + ++KD
 Sbjct: 62 ASELNRYMRGHELICQLTNDKSIPKRRRQTIDTLKKKIEEISLLIELDTENKPYQDIKD 121

5 Query: 513 ELVGDIQAQLDIELTNLQDKNTTLNKM AEVVVNLSQDNQDTKQLAKYEC SKMNL SQNV TIG 572
 ++V D+AQLD+ +T LQD LNK+AEV++NL +++ + ++LA+Y+ +KMNL+ + I
 Sbjct: 122 DIVKDMAQLDLTITELQDHIAHLNKVAEVLNLNNDIENRRRLARYDYAKMNLTA AIKIE 181

10 Query: 573 QIESEIEMIQNQLDNKIEEYENAVRKLDEYVRVLNMDK 610
 ++E EIE QN+L+ I+EYE VR+L+++ +L+ K
 Sbjct: 182 EVEKEIETSQNELNISIDEYEYLVRRLEKFG EILSDSK 219

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1977

A DNA sequence (GBSx2086) was identified in *S.agalactiae* <SEQ ID 6119> which encodes the amino acid sequence <SEQ ID 6120>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4006(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 53/115 (46%), Positives = 77/115 (66%), Gaps = 2/115 (1%)

30 Query: 5 VREIRKEVNFSEIEYQQIQNFMEQEGYEQFSPFARGKLLKIDHOPSQQLEBEWIKYLQHQK 64
 +R IRK+ + E +QI + M ++G + FS F R LL D Q +Q+E+W + QK
 Sbjct: 5 IRSIRKQFRLTETETEEKQILDLMREKGDNDNFDFLRKSLLLSDGQ--KQMEKWFNLWKKQK 62

35 Query: 65 VEQIYRDVHEILVLAKLSQSVTMEHLEIILTCTKDLMKIEVVTIPLSYSFKDKYM 119
 +EQI RDVHE+ ++AK + VT EH+ I+LTCI++L+KE+E T PLS F +KYM
 Sbjct: 63 LEQISRDVHEVFIIAKTNHQVTHEHVSILLTCTQELIKEVEKTGPLSEDFCNKYM 117

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1978

A DNA sequence (GBSx2087) was identified in *S.agalactiae* <SEQ ID 6121> which encodes the amino acid sequence <SEQ ID 6122>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

45 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]

-2224-

Identities = 176/413 (42%), Positives = 243/413 (58%), Gaps = 18/413 (4%)

Query: 26 MMFKVEAVGPPERCPECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
 M F+V+ V P C ECG + + R+ DLPI KRV L + RRRY CR C +T
 Sbjct: 1 MHFQVD-VPDPIACECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVRRRYTCCRACKTT 59

Query: 85 IS-----VDEKRSMTKRLLSIQEQSMSKTFVEVAESVGVEKTIRNVFKDYVALKERE 138
 VD R MT RL + ++++S + + VA G+DEKT+R++F R
 Sbjct: 60 FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

Query: 139 YQFETPKWLGIDEIHIIRPRLVLTNIERTTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
 ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+ R ++ V L ++ DR +E V
 Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIERTLLDLLATRRQDVVNYLMKLKDRQKVEIV 178

Query: 199 TMDMWPKYKDAVNTILPQAKVVDKHFVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
 +MDMW PY+ AV +LPQA++VVDKHFVVRMAN AL+ VRK L+ + + RTL +R
 Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKHFVVRMANDALERVKGLRKELKPSQSRTLKGDRK 238

Query: 259 ILLKRKHDNLNERESFLDDTLWGLPALKEAYELKEEFYWIWDTDPDEGHLRYSQWRHRC 318
 ILLKR H++++RE +++TW G P L AYE KE FY IWD + +W
 Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFPQLLAAYEHKERFYGIWDATTRLQAEALDEW-IAT 297

Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
 + K+ + DLVRV NW E YF D +TNAYTESIN + + R GRGYSF+ +
 Sbjct: 298 IPKGQKEVWSDLVRVGNWREETMTYFETDMPVTNAYTESINRLAKDKNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKKRKPFRFNSSAFNKAMLYDTFNWYEVNDHDITDNLGVDFSTL 429
 RA++L+ K HKK+ P S F K + Y + D N GVD ST+
 Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYKTI-----GYGLPDFAEELNYGVLDLSTI 404

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1979

A DNA sequence (GBSx2088) was identified in *S.agalactiae* <SEQ ID 6123> which encodes the amino acid sequence <SEQ ID 6124>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 412/546 (75%), Positives = 484/546 (88%)

Query: 1 MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
 M K++V+++ GMTCTGCE+HV ALE +GA IE +RRGEAVFELP+ + VE+A KAI +
 Sbjct: 1 MKKYRVDVQGMTCTGCEHVAVALENMGATGIEVDFFRGEAVFELPNALGVETAKKAISD 60

Query: 61 ANYQAGEIEEVSSLENVALINEDNYDLLIIIGSGAAAFSSAIKAIYGAQVGMIERGTG 120
 A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
 Sbjct: 61 AKYQPGKAEEVQSQEMVQLGNEGDYDIIIGSGGAAAFSSAIEAVKYGAQVAMIERGTIG 120

Query: 121 TCVNIGCVPSKTLRLRAGEINHLSKDNPFIGLQTSAGEVDLASLITQKDKLVSELNRNQKYM 180
 TCVNIGCVPSKTLRLRAGEINHLSKDNPF+GL TSAGEVDLA LI QK++LV+ELRN KY+
 Sbjct: 121 TCVNIGCVPSKTLRLRAGEINHLSKDNPFVGLHTSAGEVDLAPLIKQKNELVTELRNSKYV 180

-2225-

Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
 DLID+Y F+LI+GEAKFVD TVEVNG +SAKRFLIATGASP+ P I GL ++DYL/TST
 Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVVEVNGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQSRERLLKEYDPEISESVEKAL 300
 +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QSRERLLKEYDPEISESVEK+L
 Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQSRERLLKEYDPEISESVEKSL 300

Query: 301 IEQGINLVKGATFERVEQSSEIKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLAAAG 360
 +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
 Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGKNNELINDFGQTSNEKIYAAGDVTLGPOFVYVAAYEGGIITDNAIGGLNKKIDLS 420
 VE G EI+I+D+ +T+N +IYAAGDVTLGPOFVYVAAY+GG+ NAIGGLNKK++L
 Sbjct: 361 VEIGSRGEIIDDYSRTTNTRIYAAGDVTLGPOFVYVAAYQGGVAAPNAIGGLNKKLNLE 420

Query: 421 VVPAVTFTNPTVATVGLTEEQAKEKGVDKTSVLPPLDAVPRAIVNRETTGVFKLVADAET 480
 VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPPLDAVPRA+VNRETTGVFKLVAD++T
 Sbjct: 421 VVPGVTFTPAPAIATVGLTEQQAKENGVEVKTSLVLPPLDAVPRALVNRETTGVFKLVADSKT 480

Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDLTETLAPYLTMAEGLKLVALTFDKDISK 540
 +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
 Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

Query: 541 LSCCAG 546
 LSCCAG
 Sbjct: 541 LSCCAG 546

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1980

A DNA sequence (GBSx2089) was identified in *S.agalactiae* <SEQ ID 6125> which encodes the amino acid sequence <SEQ ID 6126>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ> .

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA83973 GB:AF138877 mercury resistance operon negative
 regulator MerR1 [Bacillus sp. RC607]
 Identities = 83/129 (64%), Positives = 104/129 (80%)

Query: 1 MIYRISEFADKCGVNKETIRYYERKNLLQEPHRTAGYRIYSYDDVKRVGFIKRIQEFGEF 60
 M +RI E ADKCGVNKETIRYYER L+ EP RTE GYR+YS V R+ FIKR+QE GF
 Sbjct: 1 MKFRIGELADKCGVNKETIRYYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

Query: 61 SLSEIYKLLGVVDKDEVRCQDMFEFVSKQKEVQKQIEDLKRIETMLDDLKQRCPEKKL 120
 +L+EI KLLGVVD+DE +C+DM++F K +++Q++IEDLKRIE ML DLK+RCP+ K +
 Sbjct: 61 TLNEIDKLLGVVDREAKCRDMYDFTILKIEDIQRKIEDLKRIERMLMDLKERCPCENKDI 120

Query: 121 HSCPIIETL 129
 + CPIIETL
 Sbjct: 121 YECPIIETL 129

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1981

A DNA sequence (GBSx2090) was identified in *S.galactiae* <SEQ ID 6127> which encodes the amino acid sequence <SEO ID 6128>. Analysis of this protein sequence reveals the following:

Possible site: 44

```
>>> Seems to have no N-terminal signal sequence
```

INTEGRAL Likelihood = -7.86 Transmembrane 80 - 96 (78 - 100)

----- Final Results -----

bacterial membrane --- Certainty=0.4142(Affirmative) < succ>

```

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8943> and protein <SEQ ID 8944> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: -13.52

GvH: Signal Score (-7.5): -6.14

Possible site: 44

```
>>> Seems to have no N-terminal signal sequence
```

```
ALOM program      count: 1 value: -7.86 threshold: 0.0
```

| | | | |
|----------|--------------------|---------------|---------------------|
| INTEGRAL | Likelihood = -7.86 | Transmembrane | 80 - 96 (78 - 100) |
|----------|--------------------|---------------|---------------------|

| | | | |
|------------|--------------|------|-----|
| PERIPHERAL | Likelihood = | 1.80 | 136 |
|------------|--------------|------|-----|

modified ALOM score: 2.07

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4142(Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF02021(439 - 666 of 1080)

GP|451734|gb|AAA18975.1||U05143 (9 - 46 of 46) envelope glycoprotein {Simian immunodeficiency virus} GP|451744|gb|AAA18980.1||U05148 envelope glycoprotein {Simian immunodeficiency virus}

```
%Match = 3.2
```

```
%Identity = 38.5    %Similarity = 64.1
```

Matches = 15 Mismatches = 13 Conservative Sub.s = 10

336 366 396 426 456 486 516 546

RTIPVOFKGCDDYYNENVGYPLSRINLEHYLTTEGGVLYFVVYSKDVSP'TV'TYASLTPKVIKNVLPASDKKKRIKKKEDI FL

[illegible]

WGLTGNAGTTTPATTTTTTTPRVVENVINESN-----

10 20 30

-2227-

```

576      606      636      666      696      726      756      786
LFWMAIIAKLLILPYPALQTSYKSRPCLRRSSLRKLTQIPFSIVTKVGNTNMKSITAFLOVKAYILPCLAKGPARIMV*W
      ||:::|  |||
-----PCIKDNSCAGLEQEP
                        40

```

SEQ ID 8944 (GBS415) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 3; MW 21.2kDa).

Example 1982

A DNA sequence (GBSx2092) was identified in *S.agalactiae* <SEQ ID 6129> which encodes the amino acid sequence <SEQ ID 6130>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.3402(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1983

A DNA sequence (GBSx2093) was identified in *S.agalactiae* <SEQ ID 6131> which encodes the amino acid sequence <SEQ ID 6132>. This protein is predicted to be ATPase. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

```

```

INTEGRAL    Likelihood = -10.08    Transmembrane  324 - 340 ( 317 - 343)
INTEGRAL    Likelihood = -5.73     Transmembrane  662 - 678 ( 660 - 690)
INTEGRAL    Likelihood = -5.41     Transmembrane  350 - 366 ( 346 - 378)
INTEGRAL    Likelihood = -3.40     Transmembrane   94 - 110 (  93 - 110)
INTEGRAL    Likelihood = -2.87     Transmembrane  681 - 697 ( 680 - 699)
INTEGRAL    Likelihood = -1.38     Transmembrane  148 - 164 ( 148 - 164)

```

```

----- Final Results -----

```

```

bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA22858 GB:M90750 cadmium-efflux ATPase [Bacillus firmus]
Identities = 486/725 (67%), Positives = 584/725 (80%), Gaps = 18/725 (2%)

```

```

Query: 1  MSRGKAKQSEKEMKAYRVQGFTCTNCAAFENNPKELPGVQDAKVNFGASKVYVKGTTTI 60
MS KA SE+EMKAYRVQGFTC NCA FE NVK+L GV+DAKVNFGASK+ V G TI
Sbjct: 1  MSDQKAITSEQEMKAYRVQGFTCANCAGKFEKNVKQLSGVEDAKVNFGASKIAVYGNATI 60

```

```

Query: 61  EELEKAGAFENLKIRDEKEQRVGGE-----PFWKQKENIKVYISALLLVVSWFL 109
EELEKAGAFENLK+ EK R + PF+K K + +Y S LL+ +
Sbjct: 61  EELEKAGAFENLKVTPEKSARQASQEVKEDTKEDKVPFYK-KHSTLLYAS-LLITFGYLS 118

```

```

Query: 110 GEQYGEHVLPTIGYAASILIGGYSLFIKGLKNLRRNLNFDMMTILMTTIAIGAAIIGEWGE 169
GEE+++ T+ + AS+ IGG SLF GL+NL R FDM TLMT+A+IG AIIGEW E

```

-2228-

5 Sbjct: 119 SYVNGEENIVTTLFLASMFIGGLSLFKVGLQNLRRFEFDMKTLMTVAVIGGAIIGEWAE 178
 Query: 170 GATVVILFAISEALERYSMDKARQSIESLMDIAPKEALIRRGNEEMMIHVDEIQVGDIMI 229
 A VVILFAISEALER+SMD+ARQSI SLMDIAPKEAL++R +E+MIHVD+I VGDIMI
 Sbjct: 179 VAIVVILFAISEALERFSMDRARQSIRSLMDIAPKEALVKRNGQEIMIHVDDIAVGDIMI 238
 10 Query: 230 VKPGQKLAMDGIVVKGSTLNQAAITGESVPVTKITNDEVFAGTLNNEEGLLEVKVKRVE 289
 VKPGQK+AMDG+VV G S +NQ AITGESVPV K ++EVFAGTLNNEEGLLEV++TK VE
 Sbjct: 239 VKPGQKIAMDGVVSGYSAVNQTAITGESVPVEKTVDNFVAGTLNNEEGLLEVETKLV 298
 Query: 290 DTTLSKIIHLVEEAQAERAPSQAFVDKFAKYTTPAIVILALLIAVPPL-FGGDWSQWIY 348
 DTT+SKIIHLVEEAQ ERAPSQAFVDKFAKYTTP I+I+A L+A+VPPL F G W WIY
 Sbjct: 299 DTTISKIIHLVEEAQGERAPSQAFVDKFAKYTTPIIIMIIATLVAIVPPLFFDGSWETWIY 358
 15 Query: 349 QGLAVLVVGCPCALVSTPFAVVTAGNAAKNGVLIKGGIHLAAGHLKAIAFDKTGTLT 408
 QGLAVLVVGCPCALV+STP+++V+AIGNAAK GVL+KGG++LE G LKAIAFDKTGTLT
 Sbjct: 359 QGLAVLVVGCPCALVISTPISIVSAIGNAAKGVLVKGGVYLEEMGALKAIADFDTGTLT 418
 20 Query: 409 KGIPAVTD--IVTYGRNENELITITSAIEKGSQHPLASAIMRKAEENGLKFNEVTVEDFQ 466
 KG+PAVTD ++ NE EL++I +A+E SQHPLASAIM+KAE + +++V VEDF
 Sbjct: 419 KGVPAVTDYNVLNKQINEKELLSIITALEYRSQHPLASAIMKAEENITYSDVQVEDFS 478
 Query: 467 SITGKGVKAKINNEMYVGSQNLFEELHSGISSDKKEKIADMQTQKTMVVLGTEKEIL 525
 SITGKG+K +N YY+GS LF+E L D ++ + +Q QGKT M++GTEKEIL
 25 Sbjct: 479 SITGKGIGIVNGTTYIIGSPKLFKELLTNDFDKLEQNVITLQNGKTAMIIGTEKEIL 538
 Query: 526 SFIADADEMRESSKEVIGKLNNGMI-ETVMLTGDNQTATAIGKQGVSDIKADLLPEDK 584
 + IAVADE+RESSKE++ KL+ +GI +T+MLTGDN+ TA AIG QGVSDI+A+L+P+DK
 Sbjct: 539 AVIAVADEVRESSKEILQKLHQLGIKKTIMLTGDNKGTANAIGGQGVSDIEAELMPQDK 598
 30 Query: 585 LNFIKELREKHQSVGMVGDGVNDAPALAASTVGVAMGGAGTDTALETADIALMSDDL SKL 644
 L+FIK+LR ++ +V MVGDGVNDAPALAASTVG+AMGGAGTDTALETAD+ALM DDL KL
 Sbjct: 599 LDFIKQLRSEYGNVAMVGDGVNDAPALAASTVGIAMGGAGTDTALETADVALMGDDLRL 658
 35 Query: 645 PYTIKLSRKALAIKQNTIFSLAIKLVALLLVMPGWLTLWIAIFADMGATLLVTLNLSRL 704
 P T+KLSRK L IIK NITF++AIK +A LLV+PGWLTLWIAI +DMGATLLV LN LRL
 Sbjct: 659 PSTVKLSRKTLNIIKANITFAIAIKFIASLLVIPGWLTLWIAILSDMGATLLVALNGLRL 718
 40 Query: 705 LKIKE 709
 +++KE
 Sbjct: 719 MRVKE 723

There is also homology to SEQ ID 3506.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 45 vaccines or diagnostics.

Example 1984

A DNA sequence (GBSx2094) was identified in *S.agalactiae* <SEQ ID 6133> which encodes the amino
 acid sequence <SEQ ID 6134>. Analysis of this protein sequence reveals the following:

50 Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0779(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2229-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1985

A DNA sequence (GBSx2095) was identified in *S.agalactiae* <SEQ ID 6135> which encodes the amino acid sequence <SEQ ID 6136>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.92    Transmembrane  123 - 139 ( 115 - 145)
      INTEGRAL    Likelihood = -6.74    Transmembrane  172 - 188 ( 167 - 190)
      INTEGRAL    Likelihood = -1.81    Transmembrane   80 - 96 ( 80 - 96)

----- Final Results -----
      bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9923> which encodes amino acid sequence <SEQ ID 9924> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 4216.

A related GBS gene <SEQ ID 8945> and protein <SEQ ID 8946> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 10
McG: Discrim Score:      -6.41
GvH: Signal Score (-7.5): -2.23
      Possible site: 58
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 3 value: -8.92 threshold: 0.0
      INTEGRAL    Likelihood = -8.92    Transmembrane  123 - 139 ( 115 - 145)
      INTEGRAL    Likelihood = -6.74    Transmembrane  172 - 188 ( 167 - 190)
      INTEGRAL    Likelihood = -1.81    Transmembrane   80 - 96 ( 80 - 96)
      PERIPHERAL  Likelihood =  2.92          46
      modified ALOM score:   2.28

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1986

A DNA sequence (GBSx2096) was identified in *S.agalactiae* <SEQ ID 6137> which encodes the amino acid sequence <SEQ ID 6138>. This protein is predicted to be histidine rich P type ATPase (HRA-1) (copB). Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -13.37   Transmembrane  318 - 334 ( 307 - 345)
      INTEGRAL    Likelihood = -5.84    Transmembrane  347 - 363 ( 335 - 364)
      INTEGRAL    Likelihood = -5.15    Transmembrane   88 - 104 ( 86 - 112)

```

-2230-

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -5.04 | Transmembrane | 651 - 667 (649 - 669) |
| INTEGRAL | Likelihood = -4.30 | Transmembrane | 156 - 172 (155 - 173) |
| INTEGRAL | Likelihood = -4.30 | Transmembrane | 669 - 685 (668 - 690) |
| INTEGRAL | Likelihood = -3.03 | Transmembrane | 62 - 78 (60 - 80) |

----- Final Results -----

bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA62113 GB:U16658 histidine rich P type ATPase [Escherichia coli]

Identities = 598/731 (81%), Positives = 651/731 (88%), Gaps = 36/731 (4%)

Query: 1 MRNNKKHSSSHSHNHGDIHSHKHDHNEHESQMDHS----- 36
MRNNK+HSSSHSHNHGD++HSHKHDHNEHESQMDHS
Sbjct: 1 MRNNKQHSSSHSHNHGDMHESKHDHNEHESQMDHSAMGHAMGGHAHHHGGMDHSHKD 60

Query: 37 -----NMDHSEMDHGAMGGHAHHHGGSFKEIFLKSPLGLIAILLITPMMDIQL 84
MD+SEMDHGAMGGHAHHHGGSFK+IFLKSPLGLIAILLITP+M IQL
Sbjct: 61 HNEMKHSQMDHSHKMDYSEMDHGAMGGHAHHHGGSFKDIFLKSPLGLIAILLITPLMGIQL 120

Query: 85 PFQIIFPYADVVAVLATILYIFGGKPFYMGAKDEFNSKAPGMSLITLGITVSYAYSIVY 144
PFQIIFPYADVVAVLATILYIFGGKPF MGAKDEFNSK PGMSLITLGITVSYAYSIVY
Sbjct: 121 PFQIIFPYADVVAVLATILYIFGGKPFMLGAKDEFNSKVPGMSLITLGITVSYAYSIVY 180

Query: 145 AVAARYVTGEHVMDFFEFTTLILIMLLGHWIEKALGEAGDAQKALAEVLPKDAHVVLE 204
AVAARYVTGE VMDFFFEFTTLILIMLLGHWIEKALGEAG+AQKALAEVLPKDAHVVLE
Sbjct: 181 AVAARYVTGEPVMDFFFEFTTLILIMLLGHWIEKALGEAGNAQKALAEVLPKDAHVVLE 240

Query: 205 DDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPIEKKTGDEVI 264
DDSIETRPV++LQ+GD+IRVQAGENVPADG I RGESRVNEALVTGESKPIEK GDEVI
Sbjct: 241 DDSIETRPVADLQVGDLIRVQAGENVPADGTIQRGESRVNEALVTGESKPIEKNPGDEVI 300

Query: 265 GGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVVAL 324
GGSTNG GVLVVEIKQTGD+SFISQVQTLISQAQSQPSRAEN+AQKVA WLFYIAV+ AL
Sbjct: 301 GGSTNGDGLVLYVEIKQTGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVIAAL 360

Query: 325 IALLIWTIIADLPTAVIFTVTALVIACPHALGLAIPLVVSRTSLGASRGLLVKNREALE 384
IAL+IW +IAD+PTAVIFTVT LVIACPHALGLAIPLV +RSTSLGASRGLLVK+R+ALE
Sbjct: 361 IALVIWMVIADVPTAVIFTVTTLVIACPHALGLAIPLVTARSTSLGASRGLLVKDRDALE 420

Query: 385 LITKADVMVLDKTGTLTTGFEKVLDDVTVLSDKYSEEEITGLLAGIEAGSSHPAQSIIVNH 444
LIT ADVMVLDKTGTLTTGFEKVLDDV + +DKY+++EI LL+GIE GSSHPAQSI+++
Sbjct: 421 LITNADVMVLDKTGTLTTGFEKVLDDVELFNDKYTKDEIVALLSGIEGGSSHPAQSIISY 480

Query: 445 AEAKGIKSVSFDSEIEVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGATLSILVENN 504
AE +GI+ VSFDSI+++SGAG+EG+ANGH YQLISQKAYG+ L MDIPKGAT+S+LVEN+
Sbjct: 481 AEQQGIRPVSVFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGATISVLVEND 540

Query: 505 EAIGAVALGDELKETSRLNIEVLKKYIEPLMATGDNEEAQGVAEVLGIQYQANQSPED 564
EAIGAVALGDELK TS++LI+ LKK I+P+MATGDNE+AAQG AE+LGI Y ANQSP+D
Sbjct: 541 EAIGAVALGDELKPTSKDLIQALKKKNKIQPIMATGDNEKAAQGAAILGIDYLANQSPQD 600

Query: 565 KYKLVESMKNQNKTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDI 624
KY+LVE +K + K VIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQ PGDI
Sbjct: 601 KYELVEKLKAEKKVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQYSPGDI 660

Query: 625 ESFIELANKTTRKMKQNLVWGAGYNFIAPIAAGLLAPIGITLGPAPGAVLMSLSTVIVA 684
SFIELA KTRKMK+NLVWGAGYNFIAPIAAG+LAPIGITL PA AVLMSLSTVIVA
Sbjct: 661 ASFIELAQKTRKMKENLVWGAGYNFIAPIAAGILAPIGITLSPAAVAVLMSLSTVIVA 720

Query: 685 INAMTLKLEPK 695
INAMTLKLEPK
Sbjct: 721 INAMTLKLEPK 731

-2231-

There is also homology to SEQ ID 3506.

A related GBS' gene <SEQ ID 8947> and protein <SEQ ID 8948> were also identified. Analysis of this protein sequence reveals the following:

```

5  Lipop: Possible site: -1  Crend: 7
   McG: Discrim Score:    -19.12
   GvH: Signal Score (-7.5): -3.71
       Possible site: 27
   >>> Seems to have no N-terminal signal sequence
   ALOM program  count: 7 value: -13.37 threshold: 0.0
10  INTEGRAL    Likelihood = -13.37  Transmembrane 291 - 307 ( 280 - 318)
   INTEGRAL    Likelihood = -5.84   Transmembrane 320 - 336 ( 308 - 337)
   INTEGRAL    Likelihood = -5.15   Transmembrane 61 - 77 ( 59 - 85)
   INTEGRAL    Likelihood = -5.04   Transmembrane 624 - 640 ( 622 - 642)
   INTEGRAL    Likelihood = -4.30   Transmembrane 129 - 145 ( 128 - 146)
15  INTEGRAL    Likelihood = -4.30   Transmembrane 642 - 658 ( 641 - 663)
   INTEGRAL    Likelihood = -3.03   Transmembrane 35 - 51 ( 33 - 53)
   PERIPHERAL  Likelihood = 0.74    103
   modified ALOM score: 3.17

20  *** Reasoning Step: 3

   ----- Final Results -----
           bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF02015(220 - 2304 of 2604)
EGAD|37454|38974(1 - 731 of 731) histidine rich P type ATPase (HRA-1) {Escherichia coli}
30 GP|643613|gb|AAA62113.1||U16658 histidine rich P type ATPase {Escherichia coli}
   PIR|JC2464|JC2464 probable copper-transporting ATPase (EC 3.6.1.-) HRA-1 -
   Enterobacteriaceae spp.
   %Match = 67.4
   %Identity = 85.9 %Similarity = 93.7
35 Matches = 598 Mismatches = 43 Conservative Sub.s = 54

162      192      222      252
PFRENYM*C*MRKF*NFKISL*YNKEELKMRNNKKHSSHSHHHNGDI-----
                                     |||||:|||||||:
40      MRNNKQHSSSHSHHHNGDMEHSHKHDHNEHMQMDHSAMGHGHCAMGGHAHHHH
                                     10      20      30      40      50

294      324      354      384      414      444      474      504
---DHSHKHDHNEHMQMDHSDHSEMDHGAMGGHAHHHHGSPFKEIFLKSPLGLIAILLITPMMDIQLPFQIIFPYADV
45  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   GMDHSHKHDHNEHMQMDHSDHSEMDHGAMGGHAHHHHGSPFKDIFLKSPLGLIAILLITPLMGIQLPFQIIFPYADV
                                     70      80      90      100     110     120     130

534      564      594      624      654      684      714      744
VAAVLATILYIFGGKPFYMGAKDEFNSKAPGMMSLITLGLITVSYAYSVAARYVTGEHVMDDFFFEFTTLILIMLLGHW
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50  VAAVLATILYIFGGKPFYMGAKDEFNSKVPKMMSLITLGLITVSYAYSVAARYVTGEPVMDFFFEFTTLILIMLLGHW
                                     150     160     170     180     190     200     210

774      804      834      864      894      924      954      984
IEMKALGEAGDAQKALAEVLPKDAHVVLEDDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55  IEMKALGEAGNAQKALAEVLPKDAHVVLEDDSIETRPVADLQVGDILRVQAGENVPADGTIQRGESRVNEALVTGESKPI
                                     230     240     250     260     270     280     290

1014     1044     1074     1104     1134     1164     1194     1224
EKKTGDEVIGGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSPSRAENVAQKVASWLFYIAVVVALIALLIWTIIAD
|| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60  EKNPGDEVIGGSTNGDGVLYVEIKQTGDKSFISQVQTLISQAQSPSRAENLAQKVAGWLFYIAVIAALIALVIWMMVIAD

```

-2232-

```

      310      320      330      340      350      360      370
1254      1284      1314      1344      1374      1404      1434      1464
5  LPTAVIFTVTALVIACPHALGLAIPVVSRSSTSLGASRGLLVKNREALELTTKADVMVLDKGTGLTTGFEKVL DVTVLSD
   :||||| ||||| ||||| ||||| :||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
VPTAVIFTVTTLVIACPHALGLAIPVLTARSTSLGASRGLLVKDRDALELTTNADVMVLDKGTGLTTGFEKVL DVELFND
      390      400      410      420      430      440      450

1494      1524      1554      1584      1614      1644      1674      1704
10 KYSEEITGLLAGIEAGSSHPAQSIIVNHAEAKGIKSVSFDSEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGA
   ||::|| ||::|| ||||| ||||| ::||| :||: ||||| ::||| :||: ||||| ||||| :| |||||
KYTKDEIVALLSGIEGSSHPAQSIISYAEQQGIRPVSFDSIDVMGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGA
      470      480      490      500      510      520      530

1734      1764      1794      1824      1854      1884      1914      1944
15 TLSILVENNEAIGAVALGDELKETSRLNIEVLKKYGIEPLMATGDNEEAAQGVAEVLGIQYQANQSPEDKYKLVESMKNQ
   |:|:||||:||||| ||::||: ||| |:|:||||:||||:|||| |:|: ||| | ||||:||||: ||| :| :
TISVLVENDEAIGAVALGDELKPTSKDLIQALKKNKIQPIMATGDNEKAAQGAAEILGIDYLANQSPQDKYELVEKLAKE
      550      560      570      580      590      600      610

1974      2004      2034      2064      2094      2124      2154      2184
20 NKTVMVMDGVNDAPSLALADVGIAIGAGTQVALDSADIILTSQSDPGDIESFIELANKTTRKMKQNLVWVGAGYNFIAIPI
   | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
GKKVIMVMDGVNDAPSLALADVGIAIGAGTQVALDSADIILTSQSPGDIASFIELAQKTTRKMKENLVWVGAGYNFIAIPI
      630      640      650      660      670      680      690

2214      2244      2274      2304      2334      2364      2394      2424
25 AAGLLAPIGITLGPFAFVAVLSLSTVIVAINAMTLKLEPK*NEAGTKKHWLV*PPSRIGSDQLVCCIRKIIDR*IFDKNR
   |||:||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAGILAPIGITLSPAAVAVLSLSTVIVAINAMTLKLEPK
      710      720      730

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 1987

A DNA sequence (GBSx2097) was identified in *S.agalactiae* <SEQ ID 6139> which encodes the amino acid sequence <SEQ ID 6140>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

```

Possible site: 59
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2197(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA40599 GB:X57326 ORF-1 [Thiobacillus ferrooxidans]
Identities = 26/65 (40%), Positives = 40/65 (61%), Gaps = 2/65 (3%)
50
Query: 1 MKQEILL--DGVK CAGCANTVQERFSAIEGVESVEVDLATKKAVLESQTEIDTETLNAAL 58
      M Q+I L G+ CA++V++ I G++S +V LAT +A + Q+ I TE L AA+
Sbjct: 1 MSQKIFLRITGMTCAHCAHSVEKALLGIHGIDSAQVSLATNQAEVFLQSSIPTALLAAV 60

55
Query: 59 AETNY 63
      + Y
Sbjct: 61 TQAGY 65

```

There is also homology to SEQ ID 3510.

-2233-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1988

A DNA sequence (GBSx2098) was identified in *S.agalactiae* <SEQ ID 6141> which encodes the amino acid sequence <SEQ ID 6142>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3220(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1989

A DNA sequence (GBSx2099) was identified in *S.agalactiae* <SEQ ID 6143> which encodes the amino acid sequence <SEQ ID 6144>. This protein is predicted to be heavy-metal transporting P-type ATPase (b0484). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.09 Transmembrane 131 - 147 ( 130 - 150)

----- Final Results -----
bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB01764 GB:U42410 heavy-metal transporting P-type ATPase
[Proteus mirabilis]
Identities = 98/153 (64%), Positives = 123/153 (80%)

Query: 2 KAVKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEVVKLQEAGKKVA 61
+A+KAL G++V MITGDNK TAKAIAKQ+GID +++EVL+ K +K+L + G KVA
Sbjct: 649 EAIALHALGLKVAMITGDNKATAKAIKQLGIDEIVAEVLPDGKVAALKQLSQKGDKVA 708

Query: 62 MVGDGINDAPALAQANVGIAVGSCTDVAIESADIVLMRNDLTAVLTITDLSHATLRNIKQ 121
VGDGINDAPALAQAVG+A+G+GTDVAIE+AD+VLM DL V+ I LS AT+RNIKQ
Sbjct: 709 FVGDGINDAPALAQADVGLAIGTGTDVVAIEAADVVLMSGDLRGVVDALALSQATIRNIKQ 768

Query: 122 NLFWAFAYNLVGIPVAMGLLYIFGGLLMSPMLA 154
NLFW FAYN + IPVA G+LY G+L+SP+ A
Sbjct: 769 NLFWTFAYNALLIPVAAGMLYPINGMLLSPIFA 801
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3505> which encodes the amino acid sequence <SEQ ID 3506>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -10.83 Transmembrane 328 - 344 ( 314 - 348)
```

-2234-

| | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -7.01 | Transmembrane | 354 - 370 (347 - 377) |
| INTEGRAL | Likelihood = -3.24 | Transmembrane | 101 - 117 (100 - 117) |
| INTEGRAL | Likelihood = -2.97 | Transmembrane | 165 - 181 (165 - 185) |
| INTEGRAL | Likelihood = -2.34 | Transmembrane | 665 - 681 (662 - 684) |
| INTEGRAL | Likelihood = -2.18 | Transmembrane | 67 - 83 (66 - 83) |
| INTEGRAL | Likelihood = -0.64 | Transmembrane | 491 - 507 (490 - 508) |
| INTEGRAL | Likelihood = -0.59 | Transmembrane | 691 - 707 (691 - 707) |
| INTEGRAL | Likelihood = -0.43 | Transmembrane | 140 - 156 (139 - 156) |

10 ----- Final Results -----

bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 92/152 (60%), Positives = 123/152 (80%)

Query: 4 VKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEVKKLQEAGKKVAMV 63
V+AL + G+ IM+TGD+ TAKAIA QVGI V+S+VLP+ KA + L+ G+KVAMV
20 Sbjct: 544 VEALHQLGIHTIMLTGDHDATAKAIASQVGITDVISQVLPDQKAGVIADLRSGRKVAMV 603

Query: 64 GDGINDAPALAQANVGIAVSGTDAIESADIVLMRNDLTAVLTITIDLSHATLRNIKQNL 123
GDGINDAPALA A++GIA+GSGTD+AIESAD++LM+ D+ ++ + LS T+R +K+NL
25 Sbjct: 604 GDGINDAPALAVADIGIAMSGTDIAIESADVILMKPMDLDLVKAMSLSRVTMRIVKENL 663

Query: 124 FWAFAYNLVGIPVAMGLLYIFGGLMSPMLAG 155
FWAF YN++ IPVAMGLL++FGG L++PMLAG
Sbjct: 664 FWAFIYNVLMIPVAMGLLLHLEGGPLLNPMLAG 695

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1990

A DNA sequence (GBSx2100) was identified in *S.agalactiae* <SEQ ID 6145> which encodes the amino acid sequence <SEQ ID 6146>. This protein is predicted to be CopY. Analysis of this protein sequence
35 reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2067(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
Identities = 63/139 (45%), Positives = 96/139 (68%)

Query: 8 TSITDAEWEVMRVWANDLVTSKTVISVLKEKMDWTESTIKTILGRLVEKGVINTEQEGR 67
TSI++AEWEVMRVWVA + +S +I++L W+ STIKT++ RL EKG L ++++GR
50 Sbjct: 2 TSISNAEWEVMRVWAKQMTSSSEIIAILSRTYCWSASTIKTLITRLSEKGYLTSQRQGR 61

Query: 68 KFIYTANIVEKEAVRDFEAEDIFNRICKKKVGNVIGSIIEDHVLSEFDDIDRLEKILEIKKS 127
K+IY++ I E+EA+ ++F+RIC K +I ++E+ ++ DI++LE +L KK+
55 Sbjct: 62 KYIYSSLISEEEALEQQVSEVFSRICVTKHQALIRHLVEETPMTLSDIEKLEALLLSKKA 121

Query: 128 FAVEEVDCCCTEGQCDCHE 146
AV EV C C GQC C+E
Sbjct: 122 NAVPEVKKNCIVGQCSCYE 140

60 There is also homology to SEQ ID 3502.

-2235-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1991

A DNA sequence (GBSx2101) was identified in *S.galactiae* <SEQ ID 6147> which encodes the amino acid sequence <SEQ ID 6148>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2829(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1992

20 A DNA sequence (GBSx2102) was identified in *S.galactiae* <SEQ ID 6149> which encodes the amino acid sequence <SEQ ID 6150>. This protein is predicted to be DS RF protein. Analysis of this protein sequence reveals the following:

Possible site: 57
>>> Seems to have a cleavable N-term signal seq.

25 INTEGRAL Likelihood =-13.21 Transmembrane 142 - 158 (136 - 169)
 INTEGRAL Likelihood = -3.45 Transmembrane 70 - 86 (66 - 88)
 INTEGRAL Likelihood = -3.13 Transmembrane 178 - 194 (176 - 195)

----- Final Results -----

30 bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAA26611 GB:L10909 putative [Staphylococcus aureus]
Identities = 98/204 (48%), Positives = 148/204 (72%), Gaps = 3/204 (1%)

Query: 4 TIISAIGVYISTSIDYLVLIILFAQLSQNKQKWHIYAGQYLGTLGVGASLVAAY-VVN 62
 TI++A VY++T IDYL++LI+LF+Q+ + + K HI+ GQY+GT +++GASL+ A VVN
Sbjct: 18 TILTATAVYVATGIDYLVILILLFSSQVKKQVK-HIWIGQYIGTAIVIGASLLVAQGVVN 76

40 Query: 63 FVPEAWMVGLLGLIPIYLGIRFAIVGEGEEEEEEIIERLEQSKANQLFWITVTLTIASG 122
 +P+ W++GLLGL+P+YLG++ I GE E+E+E I+ K NQLF T+ + +AS
Sbjct: 77 LIPQQWVIGLLGLPLYLGVKIWKGE-EDEDESSILSLFSSGKFNQLFLTMIFIVLASS 135

45 Query: 123 GDNLGIIYIPYFASLDWSQTLVLLVFAIGIIIFCELSWVLSSIPLISETIEKYQRIIVPL 182
 D+ IYIPYF +L S+ +V +VF I + + C +S+ L+S ISETIEKY+R IVP+
Sbjct: 136 ADDFSIYIPYFTTSLMSEIFIVTIVFLIMVGLCYVSYRLASFDFISETIEKYERWIVPI 195

50 Query: 183 VFIPGLGYIMYESGTIETFLNFIL 206
 VFI LG+YI++E+GT ++F+L
Sbjct: 196 VFIGLGIYTLFENGTSNALISFLL 219

-2236-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6151> which encodes the amino acid sequence <SEQ ID 6152>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have an uncleavable N-term signal seq

5 INTEGRAL Likelihood = -13.16 Transmembrane 143 - 159 (135 - 165)
 INTEGRAL Likelihood = -9.13 Transmembrane 49 - 65 (43 - 71)
 INTEGRAL Likelihood = -7.17 Transmembrane 73 - 89 (72 - 94)
 INTEGRAL Likelihood = -6.00 Transmembrane 13 - 29 (9 - 33)
 10 INTEGRAL Likelihood = -2.71 Transmembrane 180 - 196 (179 - 197)
 INTEGRAL Likelihood = -0.59 Transmembrane 112 - 128 (109 - 128)

----- Final Results -----

15 bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF42284 GB:AE002544 cadmium resistance protein [Neisseria meningitidis MC58]

20 Identities = 201/208 (96%), Positives = 205/208 (97%)

Query: 1 MRCFMIONVVTSIILYSGTAVDLLIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLL 60

MRCFMIONVVTSIILYSGTAVDLLIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLL

25 Sbjct: 1 MRCFMIONVVTSIILYSGTAVDLLIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLL 60

Query: 61 FAFVLDYIPSK EILGGLGLIPIIFLGLKVL LLDGSDGEAIAKEGLSKDNKNLIFLVAMITF 120

FAFVLDYIPSK EILGGLGLIPI LG+KVL LLDGSDGEAIAKEGL KDNKNLIFLVAMITF

30 Sbjct: 61 FAFVLDYIPSK EILGGLGLIPIILGLIKV LLDGSDGEAIAKEGLRKNKNLIFLVAMITF 120

Query: 121 ASCGADNIGVFPYF TTNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW 180

ASCGADNIGVFPYF TTNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW

Sbjct: 121 ASCGADNIGVFPYF TTNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW 180

Query: 181 FIAVVYLGLGMY ILENNSF DMLWAVLG 208

35 F+AVVYLGLG+YIL+ENNSF DMLW VLG

Sbjct: 181 FVAVVYLGLGIYILVENNSF DMLWTVLG 208

An alignment of the GAS and GBS proteins is shown below.

Identities = 71/200 (35%), Positives = 130/200 (64%), Gaps = 4/200 (2%)

40 Query: 1 MGQTIISAIGVYISTSIDYLIVLIILFAQLSQNKQKWHIYAGQYLG TGLLVGASLVAAAYV 60
 M Q ++++I +Y T++D LI+L++ FA+ K +IY GQ+LG+ L+ SL+ A+V

Sbjct: 5 MIQNVTTSIILYSGTAVDLLIILMLFFAKRKS RKDIINIYLGQFLGSVSLILLSLLFAFV 64

45 Query: 61 VNFVPEAWMVGLLGLIPIYLGIRFAIVGE GEEEEEEIIEERLEQSKANQLFWTVTLTIA 120

++++P ++GLLGLIPI+LG++ ++G+ + E + E L + N +F V ++T A

Sbjct: 65 LDYIPSK EILGGLGLIPIIFLGLKVL LLDGSDGEAIAK--EGLSKDNKNLIF-LVAMITFA 121

Query: 121 S-GGDNLGIYIPYFASLDWSQTLVLLVFAIGIIIFCELSWVLSIPLISETIEKYQR II 179

50 S G DN+G+++PYF +L+ + +V LL F + I + + L+ +P + ET+EKY R

Sbjct: 122 SCGADNIGVFPYF TTNLANLIVALLTFLVM IYLLVFS AQKLAQVPSVGETLEKYSRW 181

Query: 180 VPLVFIPLGLYIMYESGTIE 199

55 + +V++ LG+YI+ E+ + +

Sbjct: 182 IAVVYLGLGMYILENNSF D 201

SEQ ID 6150 (GBS174) was expressed in and purified from *E.coli*. The purified protein is shown in lane 7 of Figure 223.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2237-

Example 1993

A DNA sequence (GBSx2103) was identified in *S.agalactiae* <SEQ ID 6153> which encodes the amino acid sequence <SEQ ID 6154>. This protein is predicted to be Pgm. Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4324 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB96418 GB:AJ243290 phosphoglucomutase [Streptococcus thermophilus]
    Identities = 65/76 (85%), Positives = 71/76 (92%)

    Query: 1  MTYTENLQKWLD FEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
              M+YTEN QKWLD F +LP YLR EL+SMDEKTKEDAFYTNLEFGTAGMRG IGAGTNRINI
    Sbjct: 1  MSYTENYQKWLDFAELPAYLRDELVSMDEKTKEDAFYTNLEFGTAGMRGLIGAGTNRINI 60

20   Query: 61 YVVRQATEGLAKLIET 76
              YVVRQATEGLA+LI++
    Sbjct: 61 YVVRQATEGLAQLIDS 76

```

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6155> which encodes the amino acid sequence <SEQ ID 6156>. Analysis of this protein sequence reveals the following:

```

    Possible site: 53
    >>> Seems to have no N-terminal signal sequence

30   ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.4324 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

35 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 75/76 (98%), Positives = 75/76 (98%)

    Query: 1  MTYTENLQKWLD FEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
              MTYTEN QKWLD FEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI
40   Sbjct: 1  MTYTENFQKWLD FEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60

    Query: 61 YVVRQATEGLAKLIET 76
              YVVRQATEGLAKLIET
    Sbjct: 61 YVVRQATEGLAKLIET 76

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1994

50 A DNA sequence (GBSx2104) was identified in *S.agalactiae* <SEQ ID 6157> which encodes the amino acid sequence <SEQ ID 6158>. This protein is predicted to be a membrane protein. Analysis of this protein sequence reveals the following:

```

    Possible site: 53
    >>> Seems to have a cleavable N-term signal seq.

55   INTEGRAL    Likelihood = -6.21    Transmembrane    94 - 110 ( 93 - 115)
    INTEGRAL    Likelihood = -4.14    Transmembrane    172 - 188 ( 166 - 188)

```

-2238-

INTEGRAL Likelihood = -1.97 Transmembrane 130 - 146 (129 - 149)
 INTEGRAL Likelihood = -0.16 Transmembrane 62 - 78 (62 - 79)

----- Final Results -----

5 bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/185 (25%), Positives = 80/185 (42%), Gaps = 23/185 (12%)

Query: 1 MKKKNKSSNIAIIAIFFAIMLVIIHFLSSFISSFVLPVPIKPTLMHIPVIIASIAYGPRIGA 60
 MKK +I I + A+ +++ T+MHIP II I GP +G

15 Sbjct: 1 MKKSLTVRDIIVAGVLGAVAILLGVTGLGYIPVPTAAGNATIMHIPAIIIGGIMQGPVVGL 60

Query: 61 TLGALMGGISVANSSIVLLPTSILFSPFVFNENFYSLIILVPRILIGIIPYFVYKLLHN 120
 +GA+ G S N+++ L F +++++PR+ IG++ + VY +

20 Sbjct: 61 IVGAIFGISSFLNATVPL-----FKDPLVSILPRLFIGVVAWLVIYIGIRR 105

Query: 121 R---FGLAISGAIGSLTNTVFLSGIFIFFSSTYNGNIKMLLAGIISNSLAEMVIAAII 177
 + + +S IG+LTNT VL+ F + +A +N L E V+ I+

Sbjct: 106 KSEYVAVGLSAFIGTLTNTALVLA--MAVFRHYLTAGVAWTV---ITNGLPEAVVGTIV 160

25 Query: 178 VYLTV 182
 V
 Sbjct: 161 TLAVV 165

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6159> which encodes the amino acid
 30 sequence <SEQ ID 6160>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.97 Transmembrane 18 - 34 (10 - 41)
 INTEGRAL Likelihood = -7.43 Transmembrane 170 - 186 (160 - 191)
 35 INTEGRAL Likelihood = -5.63 Transmembrane 96 - 112 (94 - 117)
 INTEGRAL Likelihood = -4.67 Transmembrane 140 - 156 (131 - 158)
 INTEGRAL Likelihood = -3.66 Transmembrane 64 - 80 (63 - 84)
 INTEGRAL Likelihood = -0.22 Transmembrane 39 - 55 (39 - 55)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

>GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/193 (24%), Positives = 86/193 (44%), Gaps = 28/193 (14%)

Query: 8 RKSADISRIAIFFAIMLVIIHFVSSLVFNWPIPI---KPTLVHIPVIIASVLYGPRIGAI 64
 +KS + I I + V + P+P T++HIP II ++ GP +G I

50 Sbjct: 2 KKSALTVRDIIVAGVLGAVAILLGVTGLGYIPVPTAAGNATIMHIPAIIIGGIMQGPVVGLI 61

Query: 65 LGGLMGIISVITNTIILLPTNYLFSFVDHGTAFSLIIAIIIPRILIGITPYCYKLIPNQ 124
 +G + GI S + T+ L F +++I+PR+ IG+ + Y I +

55 Sbjct: 62 VGAIFGISSFLNATVPL-----FKDPLVSILPRLFIGVVAWLVIYIGIRR 106

Query: 125 FGLIVSGI---IGSLTNTIFVLS-GIFIFFATVFDGNIKALLTAISSNAIVEMIISAI 180
 + G+ IG+LTNT VL+ +F + T + + +N + E ++ I+

60 Sbjct: 107 SEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTA-----GVAVTVAITNGLPEAVVGTIV 160

Query: 181 TFVLIPTLSRLKR 193
 T ++ ++ R
 Sbjct: 161 TLAVVLAWKQIGR 173

-2239-

An alignment of the GAS and GBS proteins is shown below.

Identities = 121/184 (65%), Positives = 157/184 (84%)

```

5  Query: 6  KSSNIAIIAIFFAIMLVHFLSSFIFSWLVPIKPTLMHIPVIIASIAYGPRIGATLGAL 65
      KS++I+ IAIFFAIMLVHIF+SS +F+ W +PIKPTL+HIPVIIAS+ YGPRIGA LG L
      Sbjct: 9  KSADISRIAIFFAIMLVHIFVSSLVFNWIPIKPTLVHIPVIIASVLYGPRIGAILGGL 68

      Query: 66  MGGISVANSSIVLLPTSYLEFSPFVNGNFYSLIIALVPRILIGIIPYFVYKLLHNRFGLA 125
      MG ISV ++I+LLPT+YLFSPFV++G F SLIIA++PRILIGI PY+ YKL+ N+FGL
10  Sbjct: 69  MGIISVITNTIILLPTNYLFSFVDHGTFFASLIITAIIPRILIGITPYCYKLIPNQFGLI 128

      Query: 126  ISGAIGSLTNTVFLVSGIFIFSSSTYNGNIKMLLAGIISNSLAEMVIAAIIIVLTVPRI 185
      +SG IGSILTNT+FVLSGIFIFF++ ++GNIK +L IISSN++ EM+I+AI ++ +P +
15  Sbjct: 129  VSGIIGSLTNTIFVLVSGIFIFFATVFDGNIKALLTAISSNAIVEMIISAITFVLIPTL 188

      Query: 186  LNIK 189
      +K
      Sbjct: 189  SRLK 192

```

20 A related GBS gene <SEQ ID 8949> and protein <SEQ ID 8950> were also identified. Analysis of this protein sequence reveals the following:

```

25  Lipop: Possible site: -1  Crend: 5
      McG: Discrim Score: 13.42
      GvH: Signal Score (-7.5): -1.93
      Possible site: 53
      >>> Seems to have a cleavable N-term signal seq.
      ALOM program count: 2 value: -6.21 threshold: 0.0
      INTEGRAL Likelihood = -6.21 Transmembrane 94 - 110 ( 93 - 115)
      INTEGRAL Likelihood = -0.16 Transmembrane 62 - 78 ( 62 - 79)
30  PERIPHERAL Likelihood = 1.70 123
      modified ALOM score: 1.74

      *** Reasoning Step: 3

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 The protein has homology with the following sequences in the databases:

```

      ORF01561(301 - 723 of 1017)
      EGAD|38021|39600(1 - 129 of 183) hypothetical membrane protein {Bacillus acidopullulyticus}
      GP|806536|emb|CAA80247.1||Z22520 membrane protein {Bacillus acidopullulyticus}
      %Match = 7.6
45  %Identity = 29.7 %Similarity = 53.9
      Matches = 38 Mismatches = 57 Conservative Sub.s = 31

      162      192      222      252      282      312      342      372
      KKIGYQEIEPRISLLACGDTGQALADISTILKCIQEVAN*AVNLYTISSLI*GVIMKKKNKSSNIAIIAIFFAIMLVH
50  |||      :| |  :: |: :::
      MKKSLTVRDIIVAGVLGAVAILLG
      10      20

      402      432      462      492      522      552      582      612
      FLSSFIFSWLVPIKPTLMHIPVIIASIAYGPRIGATLGALMGGISVANSSIVLLPTSYLEFSPFVNGNFYSLIIALVPR
55  |:|||| || | || :| :||: | | |:: | | :::: ||
      VTRLGYIPVPTAAGNATIMHIPAIIIGGIMQGPVVGILVGAIFGISSFLNATVPL-----FKDPLVSLPR
      40      50      60      70      80

      642      663      693      723      753      783      813      843
      ILIGIIPYFVY---KLLHNRFGLAISGAIGSLTNTVFXSGIFIFSSSTYNGNIKMLLAGIISXNSLAEMVIAAIIIVLT
60  ::||:: ::|| : :| ||:|||| :| :
      LFIGVVAVWLVIYIGIRRKSEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTAGVAWTVAITNGLPEAVVGITIVTLAVVLAWK
      100      110      120      130      140      150      160

```

-2240-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1995

- 5 A DNA sequence (GBSx2105) was identified in *S.agalactiae* <SEQ ID 6161> which encodes the amino acid sequence <SEQ ID 6162>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence (or aa 1-18)

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0165(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44502 GB:U48885 DNA/pantothenate metabolism flavoprotein
[Streptococcus mutans]
Identities = 101/145 (69%), Positives = 122/145 (83%)

20 Query: 1 MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60
M K+I LAV+GSI+AYKAADL+ QLTK+GY V++ MT AA +FI PLTLQVLSKNP++ +
Sbjct: 1 MTKKILLAVSGSIAAYKAADLSHQLTKLGYHVNVMFNAAKQFIPPLTLQVLSKNPVYSN 60

25 Query: 61 VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETPKLI 120
VM E +P++INHI LAK+ DLF++ PASANT+AHLA+GFADNIVTSVALA+P E PK A
Sbjct: 61 VMKEDDPQVINHIALAKQADLFLPPASANTLAHLAHGFADNIVTSVALALPLEVPKFFA 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGY 145
PAMNTKMY N ITQ NI +LKK GY

30 Sbjct: 121 PAMNTKMYENPITQSNITLLKKFGY 145

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6163> which encodes the amino acid sequence <SEQ ID 6164>. Analysis of this protein sequence reveals the following:

Possible site: 18

35 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0076(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 146/178 (82%), Positives = 155/178 (87%)

45 Query: 1 MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60
M K ITLAV+GSISAYKAADLTSQLTKIGYDVHIIMTQAAT+FITPLTLQVLSKN IHL
Sbjct: 1 MTKHITLAVSGSISAYKAADLTSQLTKIGYDVHIIMTQAATQFITPLTLQVLSKNAIHLD 60

50 Query: 61 VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETPKLI 120
VMDEH+PK+INHIELAKRTDLFIVAPASANTIAHLAYGFADN+VTSVALA+P TPKLI
Sbjct: 61 VMDEHDPKVINHIELAKRTDLFIVAPASANTIAHLAYGFADNLVTSVALALPATTPKLI 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGYQEIEPRISLLACGDTGQALADISTILKCIQEV 178
PAMNTKMY N ITQ NI L IG+ EI P+ SLLACGD G GALADI IL I +

55 Sbjct: 121 PAMNTKMYQNPITQENIKRLSTIGFTEIPPKSSLLACGDKGPGALADIDVILATIDTI 178

SEQ ID 6162 (GBS236) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 5; MW 21.6kDa).

-2241-

Purified GBS236-GST is shown in Figure 208 (lane 6) and in Figure 225 (lanes 4-5).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1996

- 5 A DNA sequence (GBSx2106) was identified in *S.agalactiae* <SEQ ID 6165> which encodes the amino acid sequence <SEQ ID 6166>. This protein is predicted to be pantothenate metabolism flavoprotein homolog (dfp). Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2325(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

15 A related GBS nucleic acid sequence <SEQ ID 9835> which encodes amino acid sequence <SEQ ID 9836> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG39941 GB:AF301375 MTW1216 [Methanothermobacter wolfeii
      prophage psiM100]
      Identities = 71/229 (31%), Positives = 117/229 (51%), Gaps = 27/229 (11%)

Query: 6  MKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESATNL 65
      +++L++ GGT E ID VR ITN ++G +G +A + +G VTLV V + + L
Sbjct: 172 LRVLSVLSGGTLEPIDPVRVITNRRSSGRMGLAVAREAYIQGADVTLVA--GTVSVDIPSQ 229

Query: 66 STFEIEDVDSLITKPLVKEHDILIHSMASVSDYTPVYMADEFKVKSSDHLDTFLRKDNH 125
      T E + + + L+ EHD+ + + AVSD+ PVY
Sbjct: 230 RTVRAETAHEMAEAVAELIGEHDVFSAAAVSDFRPVYS----- 268

Query: 126 EGKISSESEYQVLFLKKTTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIKNK 185
      E KISS+SE L LK PK+I + ++ NP+ +VGFK V++E L AR + +
Sbjct: 269 EEKISSDSEI-TLRLKPNPKIIRMARETNPEAFIVGFKAHGVSEELIAAARKQIEDSV 327

Query: 186 ATFILANDL-IDITSKHIIAYLLDHDNVYKATT--KEDIAQLIYEKVKK 231
      A ++AND+ ++ + ++ + V + T KE++A LI ++ K
Sbjct: 328 ADMVVANDVSVEGFGSENNRAIIVSEGVTELPMTKKEELAGLIIGEIMK 376
```

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6167> which encodes the amino acid sequence <SEQ ID 6168>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1737(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 142/230 (61%), Positives = 170/230 (73%)

Query: 4  MAMKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESAT 63
      M MK++ITSGGTTE ID VR ITNH+TG LGK+I E++L+ H VTLVTTK A KP
Sbjct: 1  MTMKLIITSGGTTEPIDAVRGITNHSTGQLGKLITERFLQYHHDVTLVTTKTATKPLPNK 60
```

-2242-

Query: 64 NLSTFEIEDVDSLITLKLPLVKEHDILIHSMVSDYTPVYMADFEKVKSSDHLDTFLRKD 123
 L E+E V+ L+ LK V HDILIHSMVSDYTPVYM D E+V +D+L+ FL +
 Sbjct: 61 RLRIIEVETVNDLMAALKDQVPHHDILIHSMVSDYTPVYMTDLEQVSQADNLNCFLEH 120

5 Query: 124 NHEGKISSESEYQVLFLLKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIK 183
 N E KISS S+YQVLFLLKTPKVIS VK+WNP I LVGFKLLVNV +E L KVAR SL K
 Sbjct: 121 NSEPKISSASDYQVLFLLKTPKVISYVKQWNPNIKLVGFKLLVNVQEEELIKVARASLAK 180

10 Query: 184 NKATFILANDLIDITSKHHIAYLLDHDNVYKATTKEDIAQLIYEKVKKYD 233
 N A +ILANDL+DI + H A L+ ++ V A TKE IA L+YE++ K+D
 Sbjct: 181 NHADYILANDLVDIQTGMHKALLISNNEVASADTKEAIADLLYERMTKHD 230

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1997

A DNA sequence (GBSx2107) was identified in *S.agalactiae* <SEQ ID 6169> which encodes the amino acid sequence <SEQ ID 6170>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence
 20 INTEGRAL Likelihood = -0.22 Transmembrane 117 - 133 (117 - 133)

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9833> which encodes amino acid sequence <SEQ ID 9834> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:BAB07541 GB:AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 94/221 (42%), Positives = 133/221 (59%), Gaps = 2/221 (0%)

35 Query: 52 AEKPFITWTEVFLREINRSNQEIIILHIWPMTKTVILGMLDRELPHLELAKKEIISRGYEPV 111
 A + F + + I +S L W TV+LG+ D LP ++ + + ++ +
 Sbjct: 27 ALQSFAYDDTLCTSIGKSQSPPTLRWVHHNTVVLGIQDSRLPQIKAGIEALKGFQHDVI 86

Query: 112 VRNFGGLAVVADEGIILNFSLVIPDVFERKLSISDGYLIMVDFIRSIFSDFYQPIEHFEVE 171
 VRN GGLAVV D GILN SLV+ + E+ SI DGY +M + I S+F D + IE E+
 40 Sbjct: 87 VRNSGGLAVVLDGILNLSVLKE--EKGFSIDDGVELMYELICSMFQDHREQIEAREIV 144

Query: 172 TSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKGRSQMISDFYKIGLGD TG 231
 SYCPG +DLSI+GKKFAG++QRR I+ G+AV IYL V G R++MI FY +
 45 Sbjct: 145 GSYCPGSDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 232 SPIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVG 272
 + YP + PE MA+LS+LL P V DV+ + L++L+Q G
 50 Sbjct: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQOHG 245

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6171> which encodes the amino acid sequence <SEQ ID 6172>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have an uncleavable N-term signal seq
 55 INTEGRAL Likelihood = -0.22 Transmembrane 95 - 111 (95 - 111)

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>

-2243-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP: BAB07541 GB: AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 97/228 (42%), Positives = 138/228 (59%), Gaps = 2/228 (0%)

10 Query: 30 ALSPFVWTEVFLKTINQEPNQLILHIWPMTRTVILGMLDRQLPYFELAKTEIGNNGYVPV 89
 AL F + + + I + + L W TV+LG+ D +LP + + + +
 Sbjet: 27 ALQSFAYDDTLCTSIGKSQSPPTLRAWVHHTTVVLGIQDSRLPQIKAGIEALKGFQHDVI 86

15 Query: 90 TRNIGGLAVVADDGILNFSLVIPDHFSEISISNAYLIMVDVIRESFSDYYQRIEYHEIK 149
 RN GGLAVV D GILN SLV+ + + SI + Y +M ++I F D+ ++IE EI
 Sbjet: 87 VRNSGGLAVVLDGILNLSVLKEE--KGFSIDGYELMYELICSMFQDHREQIEAREIV 144

20 Query: 150 NSYCPGNFDSLAIAGRKFGAGIAQRRIKKGIVVSIYLSVCGDQAARGQLIKDFYEAGTQGEV 209
 SYCPG++DLST G+KFAGI+QRRI+ G+ V IYL V G A R ++I+ FY+ G+
 Sbjet: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSAERAKMIRTFYDKAVAGQP 204

25 Query: 210 TKVNPQIDPECMATLSELLETPFTVAEVLRLRLTLRQLGFSLTEKS 257
 TK YP+I PE MA+LSELL P V++VL + +TL+Q G SL +S
 Sbjet: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQQHGASLLTES 252

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 155/275 (56%), Positives = 199/275 (72%), Gaps = 8/275 (2%)

30 Query: 32 QDLAQLPVSIFKDYVTDAQDAEKPFITWTEVFLREINRSNQEIILHIWPMTKTVILGMLDR 91
 +DLA LP+ ++ D A PF+WTEVFL+ IN+ ++ILHIWPMT+TVILGMLDR
 Sbjet: 10 RDLASLPFVYGDGKVKPGALSPFVWTEVFLKTINQEPNQLILHIWPMTRTVILGMLDR 69

35 Query: 92 ELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSLVIPDVFERKLSISDGYLIMV 151
 +LP+ ELAK EI + GY PV RN GGLAVVAD+GILNFSLVIPD F +SIS+ YLIMV
 Sbjet: 70 QLPYFELAKTEIGNNGYVPVTRNIGGLAVVADDGILNFSLVIPDHFSEISISNAYLIMV 129

40 Query: 152 DFIRSIFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGD 211
 D IR FSD+YQ IE+ E++ SYCPG FDLST G+KFAG+AQRRK GI VSIYLSVCGD
 Sbjet: 130 DVIRESFSDYYQRIEYHEIKNSYCPGNFDSLAIAGRKFGAGIAQRRIKKGIVVSIYLSVCGD 189

45 Query: 212 QKGRSQMISDFYKIGLGTGSPFIAPYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQV 271
 Q R Q+I DFY+ G + + YP +DPE MA LS+LL+ P TV +V++R+ ++L+Q+
 Sbjet: 190 QAARGQLIKDFYEAGTQGEVTKVNPQIDPECMATLSELLETPFTVAEVLRLRLTLRQL 249

Query: 272 GFN-----DRLLMIRPDLVAEFNRFQAKSMANKG 300
 GF+ D+ L+ D V + R Q + + +G
 Sbjet: 250 GFSLTEKSPDQALLTNFDDAV--YERMQLVVRKEG 282

A related GBS gene <SEQ ID 8951> and protein <SEQ ID 8952> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -16.85
 GvH: Signal Score (-7.5): -5.07
 Possible site: 49

55 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -0.22 threshold: 0.0
 INTEGRAL Likelihood = -0.22 Transmembrane 117 - 133 (117 - 133)
 PERIPHERAL Likelihood = 0.47 73
 modified ALOM score: 0.54

60 *** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

ORF01564(451 - 1116 of 1518)
EGAD|13388|BS3758(27 - 249 of 281) hypothetical 31.4 kd protein in pta 3'region {Bacillus subtilis} OMNI|NT01BS4391 hypothetical protein SP|P39648|YWFL_BACSU HYPOTHETICAL 31.4 KDA PROTEIN IN PTA 3'REGION. GP|414014|emb|CAA51646.1||X73124 ipa-90d {Bacillus subtilis} GP|2636300|emb|CAB15791.1||Z99123 alternate gene name: ipa-90d {Bacillus subtilis} PIR|S39745|S39745 ywfl protein - Bacillus subtilis
%Match = 15.8
%Identity = 40.8 %Similarity = 61.0
Matches = 91 Mismatches = 82 Conservative Sub.s = 45

SEQ ID 8952 (GBS390) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 3; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A DNA sequence (GBSx2108) was identified in *S.agalactiae* <SEQ ID 6173> which encodes the amino acid sequence <SEQ ID 6174>. This protein is predicted to be probable trimethylamine dehydrogenase (nemA). Analysis of this protein sequence reveals the following:

```

50      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2218(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

>GP:CAA83700 GB:Z33015 similar to trimethylamine DH [Mycoplasma

-2245-

capricolum]

Identities = 162/311 (52%), Positives = 219/311 (70%), Gaps = 1/311 (0%)

5 Query: 3 NVQCNLFRLPLTPNGLSLENRFVLSMPVMTNSSTSEGFVTDDDIAYAVRRAKSAPLQITGA 62
 N LF P L NG LENRFVLSMP + +T +G +TD + Y RR+ SAPLQITG
 Sbjct: 2 NKYEKLFEPFYL-NGFKLENRFVLSMPMTLSLATLDGKITDKEDYVKKRRSHSAPLQITGG 60

10 Query: 63 AYITEYQQLFEYGFVSVDKEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYV 122
 Y E+GQLFEYG S D+DIP LT+L + MK+ +LQL HAG+FS +L ++GY+
 Sbjct: 61 VYFDEFGQLFEYGISAKSDDIPSLTRLYQEMKTDNSNCVILQLAHAGKFSKTSLLKKYGYL 120

15 Query: 123 YGSPMQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFF 182
 YGPS + +P H+V EL + I +II +Y AT R I+AGF+G+EIS AQRLLIQTFF
 Sbjct: 121 YGPSYEKNHTPIEHEVLELPKEKIKQIIQDYKDATLRVIKAGFNGIEISMAQRLLIQTFF 180

20 Query: 183 STFSNQRKDEYGPQTLTNRCRLGLEVFKAQVKVIREEAESDFILGFRATPEETRGSQIGY 242
 S N+R DEY NR R LEV KA+++VI + A +FI GFRATPEET G +GY
 Sbjct: 181 SQIINKRTDEYSATNFENRSRFCLEVVKAREVIDKYAPKNFIFGFRATPEETYGDILGY 240

25 Query: 243 SIEEFMEFLEKILAIQVDYLAIASWGHDFRNTIRSEGVYKQLVNVQVIFEHFGDRVPI 302
 +IE+F++ ++KI+ I ++ YLAIASWGH++ N +RS YKQLVN+VI++ + +++PI
 Sbjct: 241 TIEDFIQLVDKIIEIGKISYLAIASWGHDIYLNKVRSTKYKQLVNVKIYDIYKNKLPI 300

Query: 303 MATGGINSASK 313
 +++GGIN+ +K
 Sbjct: 301 ISSGGINTPTK 311

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6175> which encodes the amino acid sequence <SEQ ID 6176>. Analysis of this protein sequence reveals the following:

30 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3055(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 265/390 (67%), Positives = 321/390 (81%)

40 Query: 8 LFRPLTLPLNGLSLENRFVLSMPVMTNSSTSEGFVTDDDIAYAVRRAKSAPLQITGAAYITE 67
 LF PLTLPLNG L+NRFLVLSMPVMTNSST +G+VT DD++YA+RRA SAPLQITGAAY+
 Sbjct: 8 LFEPLTLPLNGSQLDNRFVLSMPVMTNSSTKDGYVTQDDVSYALRRASAPLQITGAAYVDP 67

45 Query: 68 YGQLFEYGFVSVDKEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYVYGPS 127
 YGQLFEYGFVSVDK DI GL +LA+AMK+KGAKAVLQLTHAGRF+SH L ++G+VYGPS
 Sbjct: 68 YGQLFEYGFVSVDKADISGLKELAQAMKAKGAKAVLQLTHAGRFASHALTKYGFVYGPSY 127

50 Query: 128 MQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFFSTFSN 187
 MQL+SP PH+VK LT + I +I Y QATRRAIQAGFDGVE+SSAQRLLIQTFFSTFSN
 Sbjct: 128 MQLRSPQPHEVKPLTGQQIEELIAAYAQATRRAIQAGFDGVEVSSAQRLLIQTFFSTFSN 187

55 Query: 188 QRKDEYGPQTLTNRCRLGLEVFKAQVKVIREEAESDFILGFRATPEETRGSQIGYSIEEF 247
 +R D YG QTL NR +L L V +AVQ+VI++EA FI GFRATPEETR+ IGYSI+EF
 Sbjct: 188 KRTDSYGCQTLFNRSKLTAVLQAVQVVIKQEAPDGFIFGFRATPEETRGNIDIGYSIDEF 247

60 Query: 248 MEFLEKILAIQVDYLAIASWGHDFRNTIRSEGVYKQLVNVQVIFEHFGDRVPIMATGG 307
 ++ ++ +L +A++DYLAIASWG VFRNT+RS G Y G+ VNQV+ ++ +++P+MATGG
 Sbjct: 248 LQLMDWLVNVAKLDYLAIASWGRHVFRTVRSPPGYGRRVNVVRDYLRLNKLPMATGG 307

65 Query: 308 INSASKVFEALQAHMIGASTPLVVDPEFLQKIKAKCSDQINLRIKVSDEGLAIPKASF 367
 +N+ K EAL HA IG STP VVDPEF KIK C + I+LRI+ +DL+ LAIP+ASF
 Sbjct: 308 MNTPDKATEALAHADFIGVSTPFVVDPEFAHKIKEGCEESIHLRIRPADLKSIAIPQASF 367

Query: 368 KDIVPLMDYGESLPKEAREVFRELRSNYRE 397

-2246-

KDIVPLMDYGESLPKE+R +FR L NY+E
 Sbjct: 368 KDIVPLMDYGESLPKESRTLFRSLTHNYKE 397

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1999

A DNA sequence (GBSx2109) was identified in *S.agalactiae* <SEQ ID 6177> which encodes the amino acid sequence <SEQ ID 6178>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3748(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04594 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 121/333 (36%), Positives = 192/333 (57%), Gaps = 12/333 (3%)
 Query: 1 MKLSVLDYGLIDYGKTASDAIQETILLSQEAERLGYPFWVAEHHGVKAFSISNPELMIM 60
 MKLSVLD I YG A +A+++T L++ E LGYH+FWV+EHH + S+PE++I
 Sbjct: 1 MKLSVLDQSPIAYGSSNAKEALRQTTELAKVTEALGYHRFWVSEHHDASTLAGSSSPEVLIA 60
 Query: 61 HLANQTSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNLGTVKVSNALRS 120
 HLA TK I++GSGG+M HYS++K+AE K LE HP R+ +GLG + G + ++
 Sbjct: 61 HLAHTKKIRLGSGGVMLPHYSAYKVAENFKLLEALHPGRIDVGLGRAPGGMPIAKMALQ 120
 Query: 121 LHK---AHDYEEVLEELKSWLIDESSSKEPL---VQPTLSSFPDLYVLGSGQKSAYLAA 173
 K H Y ++++ +L D+ + P + + PD+++LGS SA +AA
 Sbjct: 121 EGKEQNIHKYPLQVKDVIGYLQDDLPDHRFHGLKATPLIDTVPDVLGSSGGSANVAA 180
 Query: 174 KLGLGFTFGVFPFMDKDLTEAKKLSSLYYHQFEEYYPNKSPNLMVAAFVVIADTSEEEAE 233
 + G GF F F++ + +A + Y F+ P VA FV+ ADT E+A+
 Sbjct: 181 ENGTFGAFA--HFINGEGGVQAVE---SYRETFQPSALFDRPQTSVAIFVICADTDEQAD 235
 Query: 234 NIAKTLDIWMLGNKDFNEFATFFPTIEEANYQLTPEQKAKIKSNRDRMIVGDPKQVKESL 293
 IA +LD+ ++ ++ P+IE A Y +P ++A+I+ NR RMIVG PK V++ L
 Sbjct: 236 QIASSLDLSLIMLENGQLSKGTPSIESALSYPSPFERARIRENRKRMIVGSPKAVRQQL 295
 Query: 294 DALVNASQAEELLILIPLVPGLDQRIKSLKLLSQ 326
 L A + EE++++ + + RI+S +LL +
 Sbjct: 296 VELARAYETEEVIVVTITHRFEDRIRSYELLGE 328

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6179> which encodes the amino acid sequence <SEQ ID 6180>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 212 - 228 (210 - 229)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/329 (52%), Positives = 241/329 (72%), Gaps = 1/329 (0%)
 Query: 1 MKLSVLDYGLIDYGKTASDAIQETILLSQEAERLGYPFWVAEHHGVKAFSISNPELMIM 60

-2247-

MK+S+LDYG+ID KT +A+ ET L+Q A++LG+H+FWVAEHH + AF+IS+PEL++M
 Sbjct: 1 MKVSILDYGVVIDKEKTPQEALLETCLAQVADKLGFRFWVAEHHNIYAFAISSPELLMM 60

Query: 61 HLANQTSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNLGTVKVSNALRS 120
 HLA+ TK I+IGSGGIMPLHYSSFK+AE + TLE HPNR+ +G+GNSLGT V AL S
 Sbjct: 61 HLADHTKQIRIGSGGIMPLHYSSFKIAEWIMTLEALHPNRIDLIGNSLGTTLVQRLSS 120

Query: 121 LHKADHYEEVLEELKSWLIDESSSKEPL-VQPTLSSFPDLYVLGSGQKSAYLAACKLGLGF 179
 +H Y +V+ EL +L + S P+ V P +++P ++ L + ++A LA +LGLG+
 Sbjct: 121 IHCKDSYSQVVTELYQYLNPDLSPLPFVFNPRGNTYPQIWTLSNSLETAEAGQLGLGY 180

Query: 180 TFGVFPFMDKDPLTEAKKLSSLYHQFEYYPNKSPNLMVAAFVVIADTSEEAEINAKTL 239
 TFG+FP++ KDP+TEAK++S+ Y F K P L++A F+V++DT E+AE +AK L
 Sbjct: 181 TFGIFPYIPKDPITEAKRVSAHYRKAFRPSKLLKIPKLILAVFIVLSDTDEKAEALAKPL 240

Query: 240 DIWMLGNKDFNEFATFPTIEEANYQLTPEQKAKIKSNRDRMIVGDPKQVKESLDALVNA 299
 DIWMLG +DFNEF T+P +EEA +Y LT +Q+ I +NR RM++G P VK+ LD L+ A
 Sbjct: 241 DIWMLGQQDFNEFKTYPDVEEARNYHLTEKQREAIANRSMVIGSPHTVKKQLDRLEA 300

Query: 300 SQAEELLLIPLVPGLDQRIKSLKLLSOLY 328
 QA+ELL IPLVP R ++L+LL+ LY
 Sbjct: 301 CQADELLAIPLVPEFANRQRTLELLADLY 329

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2000

A DNA sequence (GBSx2110) was identified in *S.agalactiae* <SEQ ID 6181> which encodes the amino acid sequence <SEQ ID 6182>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF81345 GB:AC007767 Identical to a glycine cleavage system
 H-protein precursor from Arabidopsis thaliana gb|P25855.
 It contains a glycine cleavage H-protein domain
 PF|01597. ESTs gb|R90208, gb|AI
 Identities = 30/91 (32%), Positives = 53/91 (57%), Gaps = 1/91 (1%)

Query: 18 TISLTPELQDDLGTGVGYVEFTD-DANLEVDVILNIEASKTVMAILSPLTGKVVKVNTAA 76
 TI +T QD LG V +VE + ++++ + +E+ K ILSP++G+V++VNT
 Sbjct: 59 TIGITDHAQDHLGEVVFVELPEANSSVSKEKSFGAVESVKATSEILSPISGEVIEVNTKL 118

Query: 77 SQEPTLLNSEKADENWLVLTEVDYAAFEAL 107
 ++ P L+NS ++ W++ + A EAL
 Sbjct: 119 TESPGLINSSPYEDGWMIKVKPSSPAELEAL 149

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6183> which encodes the amino acid sequence <SEQ ID 6184>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2248-

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/110 (72%), Positives = 98/110 (88%)

```

5  Query: 1  MKKIANYLLIEKNEELYTISLTPELQDDLGTGVGYVEFTDDANLEVDDVILNIEASKTVMA 60
      MKKIANYLLIEK ++ YTIS+TPELQDD+GT+GY EFTD+ +L VDD+ILN+EASKTVM+
      Sbjct: 1  MKKIANYLLIEKTDDRYTISMTPELQDDIGTIGYAEFTDNDHLAVDDIILNLEASKTVMS 60

10  Query: 61  ILSPLTGKVVKVNTAASQEP TLLNSEKADENWL VVLTEVDYAAFEALENA 110
      +LSPL G VV+ N AA+ P TLLNSEKA+ENW+VVL T+VD AAF+ALE+A
      Sbjct: 61  VLSPLAGAVVERNEAATLTPTLLNSEKAENWIVVLTDVDQAAFDALEDA 110

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 2001

A DNA sequence (GBSx2111) was identified in *S.agalactiae* <SEQ ID 6185> which encodes the amino acid sequence <SEQ ID 6186>. This protein is predicted to be LRP16 (b1045). Analysis of this protein sequence reveals the following:

```

20  Possible site: 17
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0608(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:AAF15294 GB:AF202922 LRP16 [Homo sapiens]
    Identities = 73/171 (42%), Positives = 98/171 (56%), Gaps = 13/171 (7%)

    Query: 88  DICLLQVDAIVNAANSKLLGCFIPNHHCIDNQIHTFAGSRLRLACHQLMTQQGRMEAVGQ 147
      DI L+VDAIVNAANS LLG +D IH AG L C L + + G+
    Sbjct: 78  DITKLEVDAIVNAANSLLG----GGGVDGCIHRAAGPLLTDECTRLQSK-----TGK 127

35  Query: 148 AKLTESYHLPCYVIHTVGPYKVDQKPSRIREDLLKSSYKSCQLAVRANLKTIVFPCI 207
      AK+T Y LP KYVIHTVGP + S+ E L+S Y S L L + L+++ FPCI
    Sbjct: 128 AKITGGYRLPAKYVIHTVGPIAYGEPSASQAEE--LRSCYLSSLDLLLEHRLRSVAFPCI 185

40  Query: 208 STGEFGFPNQRAAELAVQAILEWQRENQHKL-YIIFNTFTPKDQDIYQKLL 257
      STG FG+P + AAE+ + + EW +++ K+ +I F KD+DIY+ L
    Sbjct: 186 STGVFGYPCEAAAEIVLATLREWLEQHKDKVDRLIICVFLEKDEDIYRSRL 236

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6187> which encodes the amino acid sequence <SEQ ID 6188>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 25
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1992(Affirmative) < succ>
50           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 139/266 (52%), Positives = 178/266 (66%), Gaps = 6/266 (2%)

```

55  Query: 1  MPNQKQLLLAMIEYLQSEKLTVDDDL----RTTDLQTVWRGLVNQDQPQNISQEYLSLED 56
      MP+ LL MI LQ+E+LT T Q +WR L+NQ+ +S++YL+LED

```


-2249-

Sbjct: 1 MPSSFDLLGEMIGLLQTEQLTSSWACPLPNALTKRQDLWRALINQRPALPLSKDYLNLED 60

Query: 57 RYLSHWNTQKVKTIDVCHQTVYSNVFTYHGDICLLQVDAIVNAANSKLLGCFIPNHHCI 116
YL W + ++ C +T Y+++F YHGD I L VDAIVNAANS+LLGCF PNH CI

5 Sbjct: 61 AYLLDDWRASFVPVSVKDCQKTNYSLSFLYHGDIRYLAVDAIVNAANSELLGCFSPNHGCI 120

Query: 117 DNQIHTFAGSRLRLACHQLMTQQGRMEAVGQAKLTESYHLPCKYVIHTVGPYVKVDQKPS 176
DN IHTFAGSRLRLAC +MT+QGR EA+GQAKLT +YHLP Y+IHTVGP + S

10 Sbjct: 121 DNAIHTFAGSRLRLACQAIMTEQGRKEAIGQAKLTSAYHLPASYIIHTVGPRIKGGHVS 180

Query: 177 RIREDLKSSYKSLQLAVRANLKTIVFPCISTGEFGFPNQRAAELAVQAILEWQRENQH 236
IR DLL Y+S L LAV+A L ++ F ISTGEFGFP + AA++A++ +L+WQ E+

Sbjct: 181 PIRADLLARCYSRLDLAVKAGLTSALFCSISTGEFGFPKKEAAQIAIKTVLKWQAEHPE 240

15 Query: 237 K--LYIIFNTFTPKDQDIYQKLLKE 260
L IFNTFT +D+ +Y L KE

Sbjct: 241 SKTLTTIFNTFTSEDKALYDTYLQKE 266

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

Example 2002

A DNA sequence (GBSx2112) was identified in *S.agalactiae* <SEQ ID 6189> which encodes the amino acid sequence <SEQ ID 6190>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2171(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6191> which encodes the amino acid sequence <SEQ ID 6192>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2477(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/284 (76%), Positives = 250/284 (87%)

Query: 4 WKTLEKTNHSQSEILSQLIEESDAIVVGIGAGMSAADGFTYIGPRFEEAFPPDFIAKYQLL 63
W T + N +Q+E L+QLI+E+DA+VVGIGAGMSAADGFTYIG RFE AFPDFIAKYQ L

Sbjct: 4 WTTYPQKNLTQAEQLAQLIKEADALVVGIGAGMSAADGFTYIGSRFETAFFPDFIAKYQFL 63

Query: 64 DMLQASLYDFEDWEEYWFQSRFVALNYLDQPVGQAYLDLKDILAKKEYHIITTNADNAF 123
DMLQASL+DFEDW+EYWFQSRFVALNYLDQPVGQ+YLDLK+IL K+YHIITTNADNAF

Sbjct: 64 DMLQASLDFEDWQEYWFQSRFVALNYLDQPVGQSYLDLKEILGT KDYHIITTNADNAF 123

Query: 124 AVADYNLEKVFHIIQGEYGLWQCSQHCHQQTyrNDQAIRQMIAQQKMDKIPSNLIPKCPKC 183
VA Y+ +FHIQGEYGLWQCSQHCHQQTy++D IRQMIA+QK+MK+P LIP CP+C

Sbjct: 124 WVAGYDPHNIFHIIQGEYGLWQCSQHCHQQTyKDDTVIRQMIAEQKNMKVPGQLIPHCPEC 183

Query: 184 DQPFEINKRNEEKGMVEDADFHQRQRYENFLSQHNDKVLYLEIGVGHTTPQFIKHPFW 243
+ PFEINKRNEEKGMVEDADFHQ+ RYE FLS+H+ KVLYLEIGVGHTTPQFIKHPFW

-2250-

Sbjct: 184 EAPFEINKRNEEKGMVEDADPHAQKARYEAFLEHKEGKVLYLEIGVGHTTPQFIKHPFW 243

Query: 244 RFVSLNENSLFVTLNHHKHYRIPQKIRSRVQLTQHIAELIAEAK 287

+ VS N N+LFVTLNHHKHYRIP IR +S++LT+HIA+LI+ K

Sbjct: 244 KRVSENPALFVTLNHHKHYRIPLSIRRQSLTEHTIAQLISATK 287

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2003

- 10 A DNA sequence (GBSx2113) was identified in *S. agalactiae* <SEQ ID 6193> which encodes the amino acid sequence <SEQ ID 6194>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1086(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12865 GB:Z99109 similar to lipoate-protein ligase [Bacillus subtilis]
Identities = 130/331 (39%), Positives = 206/331 (61%), Gaps = 5/331 (1%)

- 25 Query: 9 NGKRITDGAIALAMQVYILQNVFLDDDLFPYYCDPKVEIGKFQNAVIETNQEYLKEHDI 68

+ + I D I LA++ Y ++++ + L Y P + I G K Q N + E N +Y++E+ I

Sbjct: 5 DNQININDPRINLAIEEYCVKHLDPQQYLLFVYNQPSIIIGKNQNTIEEINTKYVEENGI 64

Query: 69 PVVRDRTGGGAVYVDSGAVNICYLMKDHGQ-FGDFKRAYEPAIKALKTLGASSVEMRERN 127

VVRR +GGGAVY D G +N ++ KD G F +FK+ EP I+AL LG + E+ RN

- 30 Sbjct: 65 IVVRRLSGGGAVYHDLGNLNFSPITKDDGDSFHNFKFTEPVIQALHQLGVEA-ELSGRN 123

Query: 128 DLVIDGKKVSGAAMTIVNGRIYGGYSLLLDVFDFAMEKVLNPNRKKIESKGIKSVRSRVG 187

D+V+DG+K+SG A GRI+ +L+ D D + L + KIESKGIKS+RSRV

- 35 Sbjct: 124 DIVVDGRKISGNAQFATKGRIFSHGTLMFDSAIDHVVSAALKVKKDKIESKGIKSIRSRVA 183

Query: 188 DIRSHLSEDIRHITTDQFKDLMVCQLLHIDHIDQAKRYHLTEKDWAIDALADEKYKNWD 247

+I L + +TT++F+ ++ + + + + Y LTEKDW I ++ E+Y+NWD

Sbjct: 184 NISEFLDDK---MTTEEFRSHLLRHIFNTNDVGNVPEYKLTEDWETIHQISKERYQNWD 240

- 40 Query: 248 WNYGNSPQYSYHRDARFPGSTYDFHLEIEKGIITNCRIYGDFFSSKDISDIENLLIGCPM 307

WNYG SP+++ + R+P G+ D HLE++KG I +C+I+GDFF D+S+IENLL+G

Sbjct: 241 WNYGRSPKFNLNHRSKRYPVGSIDLHLEVKKGKIEDCKIFGDFFGVGDVSEIENLLVGKQY 300

Query: 308 KEELVLEKLSTLSLEDYFGQTSPEEIKAVLF 338

+ ++ + L ++L+ YFG + E+ +++

- 45 Sbjct: 301 ERSVIADVLEGVNLKHYFGNITKEDFLDIY 331

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6195> which encodes the amino acid sequence <SEQ ID 6196>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0939(Affirmative) < succ>

- 55 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 248/339 (73%), Positives = 283/339 (83%)